



Exploration, par simulation, des observations grand champ d'amas de galaxies Sunyaev-Zel'dovich: intérêt en cosmologie

Jean-Baptiste Juin

► To cite this version:

Jean-Baptiste Juin. Exploration, par simulation, des observations grand champ d'amas de galaxies Sunyaev-Zel'dovich: intérêt en cosmologie. Cosmologie et astrophysique extra-galactique [astro-ph.CO]. Université Louis Pasteur - Strasbourg I, 2005. Français. NNT: . tel-00133838

HAL Id: tel-00133838

<https://theses.hal.science/tel-00133838>

Submitted on 27 Feb 2007

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

**THÈSE de DOCTORAT de l'UNIVERSITÉ LOUIS PASTEUR,
STRASBOURG**

Spécialité :

COSMOLOGIE OBSERVATIONNELLE

présentée par

Jean-Baptiste JUIN

pour obtenir le TITRE de DOCTEUR de l'UNIVERSITÉ LOUIS PASTEUR

**Exploration, par simulation,
des observations grand champ
d'amas de galaxies Sunyaev-Zel'dovich
- Intérêt en cosmologie -**

soutenu le 21 septembre 2005 devant le jury composé de :

MM	James G. BARTLETT	Président
	Ulrich GOERLACH	Directeur de thèse
	Dominique YVON	Co-directeur de thèse
	Jean HEYVAERTS	Rapporteur interne
	James G. BARTLETT	Rapporteur externe
	François-Xavier DÉSSERT	Rapporteur externe

Remerciements

En premier lieu, je tiens à remercier le rectorat de l'académie de Strasbourg pour avoir financé ces trois années de recherche qui m'ont permis d'obtenir le grade de docteur. Ceci n'aurait pas été possible sans l'aide de M. Bing, directeur du DEA de physique subatomique de Strasbourg ; je le remercie. Je remercie aussi le Service de Physique des Particules du Dapnia pour m'avoir fourni tous les moyens de réaliser ce travail de recherche, que ce soit au niveau de l'encadrement ou des moyens matériels. C'est un laboratoire où il fait bon être en thèse. Dans ce contexte, je remercie Bertrand Vallage d'avoir bien voulu être mon parrain de thèse. Il a su me conseiller pour assurer le bon avancement de mon travail. Je remercie vivement, par ailleurs, Yannick Giraud-Héraud, pour m'avoir "hébergé" dans les locaux du PCC pendant les quatres premiers mois de ma thèse. Je le remercie aussi de m'avoir aidé lors de l'organisation de la soutenance.

Je remercie vivement mon directeur de thèse et encadrant, Dominique Yvon, pour sa disponibilité, son expérience et pour les nombreuses discussions (parfois enflammées) que nous avons dû avoir au cours de ces trois années, que ce soit sur des sujets de physique ou sur plein d'autres choses. J'espère vivement pouvoir travailler à nouveau avec lui dans l'avenir.

Je remercie Ulrich Goerlach, mon directeur de thèse, pour son aide dans la gestion de la fin de thèse, moment toujours délicat et qu'il est difficile d'aborder car la somme de travail augmente avec les derniers résultats.

Je remercie Jean Heyvaerts, François-Xavier Désert et James G. Bartlett d'avoir endossé le rôle de rapporteur. Leurs remarques m'ont permis de grandement améliorer le manuscrit et m'ont aidé à prendre du recul sur mon travail.

Je remercie vivement Alexandre Réfregier sans qui ce travail n'aurait pu, ni commencer, ni aboutir. Il a été un guide scientifique disponible et précieux par la variété de ses compétences et connaissances. Les nombreuses discussions que nous avons eu ont été riches en enseignements et en conséquences sur mon travail.

Je remercie aussi Sandrine Pires et Yassir Moudden pour m'avoir initié, en douceur, aux mystères des séparations en composantes indépendantes et des *prior* entropiques. Ils m'ont aidé à accoster au rivage de ce nouveau continent à explorer !

Je remercie par ailleurs, Jean-Pierre Pansart et Christophe Magneville pour leurs conseils et leur disponibilité. Merci aux habitués du repas de midi, moment de véritable détente : indispensable.

Pendant la période au PCC, j'ai pu rencontrer Jean-Baptiste Melin, je le remercie pour les discussions que nous avons eu alors sur "la thèse" en général et pour toutes les discussions que nous avons eu après. Je lui souhaite une excellente continuation au SPP.

Merci à mon amie Charlotte, qui m'a soutenu de tout son coeur pendant ces trois années. Je remercie surtout mes parents, Philippe et Isabelle, pour leur soutien, total et immuable.

Résumé

L'objectif de mon travail de recherche est de préparer l'analyse des données des prochaines observations grand champ d'amas de galaxies par effet Sunyaev Zel'dovitch. Pour cela, je mets en place une chaîne complète d'outils permettant d'effectuer cette étude. Ces outils permettent de mettre en évidence des points critiques des effets de sélection à prendre en compte dans les futures analyses. La chaîne d'analyse est constituée : d'une simulation de ciel observé millimétrique, d'algorithmes originaux d'extraction d'amas SZ à partir des cartes observées, d'un modèle statistique des effets de sélection de la chaîne de détection, et enfin d'outils d'analyse des catalogues des sources détectées pour contraindre les paramètres cosmologiques. Je m'intéresse au cas des expériences multi-fréquences dotées de caméra bolométrique. J'applique ces outils à une prospective pour l'expérience Olimpo.

Abstract

The goal of my Phd research is to prepare the data analysis of the near future wide-field observations of galaxy clusters detected by Sunyaev Zel'dovitch effect. I set up a complete chain of original tools to carry out this study. These tools allow me to highlight critical and important points of selection effects that has to be taken into account in future analysis. Analysis chain is composed by : a simulation of observed millimeter sky, state-of-the-art algorithms of SZ galaxy clusters extraction from observed maps, a statistical model of selection effects of the whole detection chain and, finally, tools to constrain, from detected SZ sources catalog, the cosmological parameters. I focus myself on multi-channel experiments equipped with large bolometer camera. I use these tools for a prospection on Olimpo experiment.

Table des matières

Introduction	i
I <i>Cosmologie, amas de galaxies et effet SZ</i>	1
1 Cosmologie	3
1.1 Vers le modèle du Big-Bang	5
1.1.1 Mesure du décalage vers le rouge des spectres galactiques . . .	5
1.1.2 Mesure de l'abondance des éléments légers	5
1.1.3 Observation du rayonnement à $3K$	6
1.1.4 Le modèle du Big-Bang	6
1.1.5 Le contenu de l'univers aujourd'hui	7
1.1.5.1 Les baryons lumineux	7
1.1.5.2 La matière sombre	7
1.1.5.3 L'énergie sombre	9
1.2 Description de l'univers homogène et isotrope	10
1.2.1 Le principe cosmologique	10
1.2.2 Relativité et gravitation	10
1.2.3 La métrique de Robertson-Walker	11
1.2.4 Le tenseur énergie-impulsion	11
1.2.5 La dynamique de Friedmann-Lemaître	12
1.2.6 Des quantités importantes pour la suite	12
1.2.7 Une brève histoire thermique	15
1.3 L'univers inhomogène : de l'inflation au découplage	19
1.3.1 L'inflation	19
1.3.2 Évolution du champ de densité	20
1.3.2.1 La matière noire, un fluide à pression nulle	21
1.3.2.2 Description du mélange baryons-photons	24
1.3.2.3 Echelle caractéristique de l'évolution des surdensités	26
1.3.3 Après le découplage	27
1.3.3.1 Le CMB	27
1.3.3.2 Vers la formation des grandes structures	30

2	Amas de galaxies : formation, observations SZ et cosmologie	33
2.1	Les amas : observations multi-longueurs d'onde	35
2.1.1	Premières études en optique : les galaxies	35
2.1.2	L'apport des observations en X : étude du gaz	36
2.1.3	Les observations SZ pointées	36
2.1.4	Le weak lensing : sonder la matière sombre	36
2.1.5	Un amas en quelques chiffres	37
2.2	Description individuelle des amas	39
2.2.1	Le modèle sphérique d'effondrement gravitationnel	39
2.2.2	Physique de l'état lié	43
2.2.2.1	Les profils de densité	43
2.2.2.2	Calcul du rayon viriel : $R_v(M)$	45
2.2.2.3	Calcul de la température : $T(z, M)$	45
2.2.3	Effondrement asymétrique	46
2.3	Description statistique des amas : fonction de masse	47
2.3.1	Formalisme de Press & Schechter	47
2.3.2	D'autres fonctions de masse	50
2.3.3	Calcul de la distribution en masse et <i>redshift</i> des amas	50
2.4	Observations par effet SZ thermique	53
2.5	Amas de galaxies et cosmologie	57
3	L'expérience Olimpo : contexte et instrument	59
3.1	Contexte	60
3.2	L'instrument Olimpo	63
3.2.1	Telescope	63
3.2.2	Cryostat	63
3.2.3	Bolomètres	63
3.2.4	Système de pointage	66
3.3	Comparaison avec APEX et Planck	67
II	<i>Les sondages SZ : effets de sélection et cosmologie</i>	69
4	Simulation d'observation	75
4.1	Le ciel sub-millimétrique	76
4.2	Les cartes d'effet SZ thermique	78
4.2.1	Description des modèles utilisés	78
4.2.2	Calcul des distributions théoriques simulées	80
4.2.3	Calcul de la carte	83
4.3	Les phénomènes physiques "contaminants"	86
4.3.1	Le CMB	86
4.3.2	Les galaxies infrarouges	86
4.3.3	La poussière galactique	88
4.3.4	L'atmosphère	89
4.4	Spectres d'émissions des sources astrophysiques	90
4.5	Observations multi-fréquences bolométriques	93
4.5.1	Les fréquences d'observation	93

4.5.2	Bruit de mesure	93
4.5.3	Stratégie d'observation	95
4.5.4	Lobe d'antenne et précision de pointage	95
4.6	Conclusion	97
5	Détection d'amas sur les cartes simulées	99
5.1	Introduction	100
5.2	Isoler l'effet SZ	103
5.2.1	Séparation aveugle de sources	103
5.2.2	JADE	104
5.2.3	Représentation dans une base multi-échelles	105
5.2.4	Résultats	106
5.2.4.1	Séparation en aveugle : JADE avec des données bruitées	106
5.2.4.2	Comment améliorer les choses ?	109
5.2.4.3	Conclusion	109
5.3	Filtrage des cartes	113
5.3.1	Filtrage Gaussien	113
5.3.2	Filtrage non-linéaire : ME-FDR	113
5.3.2.1	Maximum Entropy Method	114
5.3.2.2	Multiscale Entropy	115
5.3.2.3	Filtrage ME-FDR	117
5.3.3	Résultats	118
5.4	Détection des amas	123
5.4.1	Algorithme de détection de sources brillantes	123
5.4.2	Association et marquage des sources	125
5.4.3	Qualité du catalogue reconstruit	127
5.4.4	Résultats	127
5.5	Conclusion	130
6	Modélisation statistique des effets de sélection	135
6.1	Introduction : les effets de sélection	136
6.2	Modèle statistique d'Olimpo	139
6.2.1	Complétude	140
6.2.2	Erreurs photométriques	143
6.2.2.1	Estimation de Y_{obs}	144
6.2.2.2	Estimation de z_{obs}	149
6.2.3	Les contaminations	149
6.2.4	Nombre effectif de détections	151
6.3	Validité du modèle statistique	154
6.3.1	Complétude, photométrie et contaminations	154
6.3.2	Modèle statistique et confusion	157
6.4	Conclusion	163

7	Contraintes sur les paramètres cosmologiques	165
7.1	Observations et cosmologie	166
7.2	Outils statistiques d'analyse des données	168
7.2.1	Compatibilité modèle-observation : les tests d'hypothèses . . .	169
7.2.1.1	Interêt	169
7.2.1.2	L'estimateur de vraisemblance	169
7.2.1.3	Test d'hypothèse en pratique	170
7.2.1.4	Rejeter Press & Schechter	171
7.2.2	Estimation des paramètres	172
7.2.2.1	Cartes de vraisemblance	173
7.2.2.2	Calcul des niveaux de confiance par Monte-Carlo . .	173
7.2.2.3	L'estimateur de vraisemblance	175
7.3	Perspectives pour Olimpo	178
7.3.1	SZ, CMB et <i>weak-shear</i> : contraintes sur T_\star	179
7.3.2	Ω_M et σ_8	181
7.3.3	Variance des comptages et contraintes	181
7.3.4	Suivi optique partiel	184
7.3.5	Contraintes "robustes"	184
7.4	Systématiques liées au modèle statistique	187
7.4.1	Estimateur des paramètres cosmologiques	187
7.4.2	Modèle statistique et confusion	188
7.4.3	Complétude	191
7.4.4	Biais photométrique	191
7.4.5	Contamination	192
7.5	Conclusion	194
	Conclusion	195

Introduction

La cosmologie s’efforce de comprendre l’évolution de l’univers dans son ensemble à travers l’observation de son contenu. Aujourd’hui un certain nombre d’observations et de mesures complémentaires, du CMB, des supernovae, de l’abondance des éléments légers, de la vitesse d’éloignement des galaxies proportionnelle à leur distance, confirment le modèle du *Big-Bang* chaud et mettent en avant un modèle cosmologique consensuel décrit par un jeu de paramètres cosmologiques. Dans ce modèle, le contenu visible de l’univers est sous-dominant. La dynamique de l’univers est dominée par des composantes invisibles directement : la matière sombre et l’énergie sombre qui comptent respectivement pour 25 et 70% de la densité d’énergie de l’univers.

Avant l’époque du découplage entre le rayonnement et la matière, la dynamique de l’univers est dominée par le rayonnement. Par ailleurs, durant cette période, la matière sombre, qui se comporte gravitationnellement comme la matière, est le moteur de la croissance des fluctuations de densité primordiales à l’origine des amas de galaxies.

L’histoire des amas de galaxies, les structures les plus massives de l’univers (jusqu’à $10^{15} M_{\odot}$), est riche en information sur ces composantes sombres. En effet, les amas se forment sous l’action gravitationnelle de la matière sombre dans un univers dont l’expansion globale est aujourd’hui dominée par l’énergie sombre. La mesure de l’abondance des amas en fonction du temps (i. e., le redshift) permet de tracer l’évolution de l’univers à grande échelle et de mieux comprendre les composantes sombres qui dominent la dynamique à ces échelles.

Les amas sont des structures complexes. Ils sont composés principalement : de galaxies (de quelques dizaines à quelques centaines), elles-mêmes composées d’étoiles (plusieurs milliards), d’un gaz baryonique chaud ionisé, dont la température électronique est de l’ordre de 10^7 K, et de matière sombre, à l’origine, pour la plus grosse partie, du puits de potentiel gravitationnel de l’amas. Des observations d’amas proches en optique, par rayonnement X, par effet Sunyaev Zel’dovitch (SZ) et par le biais des lentilles gravitationnelles, ont mis en évidence les phénomènes physiques complexes qui ont lieu au sein des amas : chauffage du gaz par la gravitation, formation d’étoiles, phénomènes de refroidissement du gaz, collisions entre galaxies, etc. Des modèles ont, cependant, pu être mis en place pour décrire un comportement “moyen” des amas, notamment en ce qui concerne des relations entre quantités intégrées sur l’amas : masse-température ou masse-rayon par exemple, mais aussi les profils internes de densité (de matière sombre, des galaxies ou du gaz), ou plus récemment les profils de température. Cette complexité rend les amas d’autant plus important à étudier qu’il est nécessaire, pour utiliser cette sonde astrophysique, dans

un contexte cosmologique, de bien comprendre leurs mécanismes de formation. En effet on constate des dégénérescences entre les informations cosmologiques et celles provenant de la physique interne des amas.

L'effet SZ se révèle être, non-seulement une sonde puissante pour étudier la physique interne des amas au delà de leur rayon de coeur (combinée avec les observations en X, plus sensibles au centre des amas), mais aussi un outil de choix pour étudier la cosmologie via les amas de galaxies. En effet, le flux SZ, sur la ligne de visée, présente la particularité d'être indépendant du *redshift* (i.e., de la distance) de l'amas source, autorisant la détection d'amas lointains, riches en information sur la cosmologie. De plus l'effet SZ permet, de manière plus directe qu'avec les observations en X, de tracer la masse des amas.

Grâce aux performances croissantes des instruments, les expériences permettent aujourd'hui d'observer des amas par effet SZ sur une large gamme de *redshifts* et apportent d'ores et déjà une riche information sur les amas et sur la cosmologie. Dès 2006, la technologie liée, entre autre, aux caméras bolomètres, va permettre d'effectuer des sondages grand champ (plusieurs centaines de degrés carrés) d'amas par effet SZ. On s'attend, dans les années qui viennent, à observer des catalogues de plusieurs centaines, voir milliers d'amas SZ, débouchant sur un apport important à la physique des grandes structures, aux composantes sombres de l'univers et à la cosmologie. Je me suis intéressé plus particulièrement à l'instrument Olimpo dont le vol scientifique est prévu courant 2006. Entre 2006 et 2008 le télescope APEX et le satellite Planck vont effectuer, eux aussi, des observations grand-champ d'amas SZ. Ces deux prochaines années seront riches en observations SZ de seconde génération, et on peut compter sur la détection de larges catalogues d'amas.

Le travail effectué durant ma thèse est une préparation à l'analyse des données de ces prochains sondages SZ grand champ. Grâce à la mise en place d'une chaîne complète de simulation, détection et analyse des catalogues SZ, je montre l'importance de la compréhension des effets de sélection, dans ce type d'observation, pour extraire l'information scientifique des données et contraindre les paramètres cosmologiques.

Ce travail s'appuie sur une simulation réaliste de ciel millimétrique observé. Elle inclut les phénomènes astrophysiques principaux à ces fréquences : CMB, effet SZ, galaxies infrarouges et poussière galactique. Les galaxies infrarouges sont simulées, chacune avec un spectre d'émission propre ce qui est effectivement observé en pratique.

Les méthodes de détection d'amas devront permettre de traiter le cas particulier des galaxies infrarouges qui, sinon, contamineront les détections. Par ailleurs, les méthodes de détection d'amas, doivent permettre, dans un premier temps, d'extraire un maximum d'amas des cartes observées tout en limitant la présence de contaminations, mais aussi d'effectuer une photométrie fidèle et précise des amas. Nous avons mis au point une chaîne de détection et d'extraction des amas SZ originale ayant ces objectifs. La détection des amas se fait en trois étapes : 1) la séparation de l'effet SZ des autres composantes astrophysiques par l'utilisation combinée de ses signatures

particulières : spectrale et spatiale, 2) l'augmentation du rapport signal sur bruit par la mise en place d'un filtrage non-linéaire basé sur une méthode entropique, 3) l'extraction des amas en tant que sources brillantes SZ, en tenant compte de leur éventuelle superposition, et l'estimation de leur flux. Grâce à l'utilisation de plusieurs critères j'ai comparé la qualité de notre chaîne de détection à des méthodes plus classiques de détection de sources brillantes. Elle se révèle efficace et j'ai pu mettre en avant une des méthodes permettant de limiter la contamination dans les catalogues de sources détectées. Ceci se révèle un point important pour envisager de contraindre au mieux les modèles cosmologiques.

L'utilisation du catalogue des sources détectées pour étudier la cosmologie a été l'objet d'une attention particulière. Les dégénérescences entre certains paramètres des modèles de physique interne des amas et cosmologique, et le fait que, dans un premier temps, les catalogues SZ compteront peu de sources, m'ont incité à mettre en place un outil de test d'hypothèse permettant de chercher une classe de modèles cosmologiques statistiquement compatibles avec les données recueillies. Ce point, souvent laissé de côté, se révèle d'importance pour pouvoir faire confiance à l'estimation des paramètres proprement dite. Par ailleurs, pour cette tâche, et suivant la démarche précédente, j'ai mis en place un outil de calcul des contraintes, permettant, par Monte-Carlo, d'estimer les niveaux de confiances des modèles cosmologiques retenus comme compatibles avec les données. Ce travail d'analyse s'appuie sur l'utilisation d'un modèle statistique des effets de sélection de la chaîne de détection SZ complète (instrument et algorithmes d'extraction des amas).

Ce modèle, construit par Monte-Carlo, révèle la complexité des effets de sélection et joue un rôle central dans l'estimation des paramètres cosmologiques. En effet il permet de faire le lien entre le catalogue des détections extrait des données et la distribution réelle des amas dans le champ, elle-même reliée à la cosmologie *via* les modèles de formation de structures. J'étudie et modélise, entre autre, l'efficacité de détection réelle d'un amas en fonction de son flux et de son redshift, allant plus loin qu'un simple seuil de détection en flux (hypothèse souvent faite). La modélisation, en final, prend en compte cette "complétude" du sondage, les erreurs photométriques, les contaminations et le nombre effectif de détections attendu pour une observation. Ce dernier point est critique. Je montre que la variance des comptages est plus compliquée que ce qui est généralement admis.

Ainsi mon travail de thèse a consisté en la mise au point d'un ensemble cohérent d'outils permettant d'étudier les effets de sélection (astrophysiques, instrumentaux ou algorithmiques) liés aux observations SZ grand champ depuis l'instrument jusqu'aux contraintes. Ce travail constitue une étape essentielle aux analyses des prochaines données de ce type de sondage.

Le manuscrit se décompose en deux parties. La première présente le contexte physique de mon travail : le modèle cosmologique du big-bang (**chapitre 1**), les modèles décrivant la physique interne des amas et l'effet SZ thermique (**chapitre 2**) et enfin, le contexte instrumental actuel des observations grand-champ d'amas de galaxies par effet SZ (**chapitre 3**). La seconde partie présente mon travail de thèse

proprement dit. Tout d'abord la simulation de ciel millimétrique observé (**chapitre 4**), la méthode de détection d'amas SZ dans le champ observé (**chapitre 5**), les effets de sélection des observations SZ grand-champ et le modèle qui a été mis en place (**chapitre 6**), et pour finir, les outils de calcul des contraintes des modèles cosmologiques et l'impact des effets de sélection sur ceux-ci (**chapitre 7**). Ce travail fait l'objet de deux publications soumises à la revue A&A, dont les résumés sont donnés en annexe.

Première partie

Cosmologie, amas de galaxies et effet SZ

Chapitre 1

Cosmologie

Sommaire

1.1	Vers le modèle du Big-Bang	5
1.1.1	Mesure du décalage vers le rouge des spectres galactiques	5
1.1.2	Mesure de l'abondance des éléments légers	5
1.1.3	Observation du rayonnement à $3K$	6
1.1.4	Le modèle du Big-Bang	6
1.1.5	Le contenu de l'univers aujourd'hui	7
1.1.5.1	Les baryons lumineux	7
1.1.5.2	La matière sombre	7
1.1.5.3	L'énergie sombre	9
1.2	Description de l'univers homogène et isotrope	10
1.2.1	Le principe cosmologique	10
1.2.2	Relativité et gravitation	10
1.2.3	La métrique de Robertson-Walker	11
1.2.4	Le tenseur énergie-impulsion	11
1.2.5	La dynamique de Friedmann-Lemaître	12
1.2.6	Des quantités importantes pour la suite	12
1.2.7	Une brève histoire thermique	15
1.3	L'univers inhomogène : de l'inflation au découplage . . .	19
1.3.1	L'inflation	19
1.3.2	Évolution du champ de densité	20
1.3.2.1	La matière noire, un fluide à pression nulle . . .	21
1.3.2.2	Description du mélange baryons-photons	24
1.3.2.3	Echelle caractéristique de l'évolution des surdensités	26
1.3.3	Après le découplage	27
1.3.3.1	Le CMB	27
1.3.3.2	Vers la formation des grandes structures	30

L'objet de ce chapitre est d'introduire le contexte général de l'évolution de la matière dans l'univers. Je me suis attaché à décrire l'évolution de la matière à travers les différentes ères de l'univers afin de parvenir, en fin de chapitre, au début de la période de la formation des amas. Le chapitre 2 traitant de la description plus détaillée de leur évolution. Ce chapitre n'a pas pour objectif de décrire en détail les différents phénomènes physiques abordés, de nombreux livres et revue de cosmologie remplissent ce rôle, notamment le livre de Peebles (1980) ou celui, plus récent de J. Rich : Rich (2002). Ainsi le présent chapitre aborde les points suivants :

1. les observations cosmologiques fondamentales, bases de la cosmologie moderne, aboutissant au modèle du Big-Bang,
2. une description du contenu de l'univers tel qu'on le connaît actuellement,
3. une description de l'évolution moyenne de l'univers dans le cadre du Big-Bang,
4. une description plus précise de cette évolution en terme de fluctuations de densité.

1.1 Vers le modèle du Big-Bang

Je présente dans cette section les trois observations majeures qui posent les bases de la cosmologie moderne. Elles concernent des domaines de physique différents et ont été le fruit de prouesses technologiques et scientifiques. Ces trois observations ont permis d'établir le modèle du Big-Bang.

1.1.1 Mesure du décalage vers le rouge des spectres galactiques

En 1929 Hubble montre que le décalage vers le rouge du spectre des galaxies lointaines est proportionnel à leur distance. Ce décalage ayant déjà été constaté, mais non expliqué, par Slipher en 1918. Hubble a utilisé les étoiles de type "Céphéide" pour estimer la distance des galaxies qu'il a observé. Ces étoiles ont une relation période-luminosité bien connue. Le *redshift* z des galaxies est par définition :

$$z = \frac{\lambda_o - \lambda_e}{\lambda_e}$$

où λ_o est la longueur d'onde observée par l'astronome et λ_e celle à laquelle le rayonnement a été émis. En interprétant ce décalage vers le rouge comme un effet Doppler, Hubble a déduit une vitesse d'éloignement des galaxies lointaine $v = cz$. L'observation des Céphéides va lui permettre de relier cette vitesse à la distance d entre nous et ces galaxies, et d'en déduire la relation :

$$\frac{v}{d} = H_0$$

où H_0 est la constante de Hubble. Aujourd'hui cette relation n'est plus interprétée comme un "éloignement" des galaxies les unes des autres mais comme une expansion de l'univers. Toutes les galaxies semblent s'éloigner les unes des autres de manière isotrope.

De nombreuses expériences ont été mise en place pour mesurer cette valeur de H_0 . Cette mesure est difficile car les vitesses apparentes d'éloignement des galaxies sont contaminées par leur vitesse particulière. Par ailleurs les mesures de distance sont des mesures difficiles et contribuent aux incertitudes sur la valeur de H_0 .

Aujourd'hui la communauté scientifique s'accorde sur la valeur donnée par le satellite Hubble (Freedman et al. 2001) :

$$h = \frac{H_0}{100 \text{ km/s/Mpc}} = 0.72 \pm 0.08$$

1.1.2 Mesure de l'abondance des éléments légers

Après la découverte de Hubble et l'interprétation du décalage vers le rouge des galaxies, les astrophysiciens ont supposé qu'en remontant le temps l'univers doit devenir de plus en plus chaud et dense. Suivant cette hypothèse ils ont observé

que le rapport des abondances des éléments ${}^4\text{He}$ et H dans différents corps astrophysiques semble universel. Ce qui implique un lien à l'échelle cosmologique. C'est Gamow, en 1930, qui suggère que les réactions nucléaires dans un univers initialement dense et chaud pourrait conduire à de telles abondances à travers le mécanisme de nucléosynthèse primordiale.

En suivant cette voie on peut montrer que les éléments légers (Hydrogène, Deutérium, Hélium 3, Hélium 4, Lithium) ont été formés dans l'univers primordial. Les éléments plus lourds sont formés au coeur des étoiles par des réactions complexes et tardives. Les abondances primordiales D/H , ${}^4\text{He}/H$ ont été estimées dans différents lieux, à différentes échelles : système solaire, gaz interstellaire, planètes gazeuses, etc. La meilleure contrainte (Tytler et al., 2000) donne pour η , le rapport des abondances de deutérium et d'hydrogène :

$$D/H : \eta = 5.1 \pm 0.5 \cdot 10^{-5}$$

La cohérence des résultats en différents lieux astrophysiques confirme l'hypothèse de Gamow d'un univers primordial dense et chaud.

1.1.3 Observation du rayonnement à 3K

Poursuivant cette voie, Gamow, Alpher et Herman, à la fin des années 40, aboutissent à la prédiction de l'existence d'un rayonnement de corps noir remplissant tout l'univers et dont la température peut être directement reliée à la densité de baryons dans l'univers primordial. C'est en 1965 que cette prédiction sera confirmée par l'observation, par Penzias et Wilson, de ce rayonnement. Ils confirmeront son caractère isotrope et donneront la température du corps noir correspondant de $3.5K$. Cette observation est replacée dans le contexte cosmologique la même année par Dicke, Peebles, Roll et Wilkinson (Dicke et al., 1965).

Cette observation est l'un des plus grand succès de la théorie d'un univers primordial dense et chaud en expansion (modèle du *Big-bang*). Depuis cette observation de nombreuses expériences se sont succédées pour étudier ce rayonnement, appelé dans la suite CMB (pour *Cosmic Microwave Background*). Le spectre du CMB est mesuré en 1989 par FIRAS (*Far Infrared Absolute Spectrophotometer*) placé sur le satellite COBE. Les astrophysiciens en déduiront la température du corps noir (Kashlinsky et al., 1999) :

$$T_{CMB} = 2.725 \pm 0.001K$$

C'est la preuve éclatante confirmant le modèle du Big-Bang. En 1992, COBE emporte une nouvelle expérience d'observation du CMB : *DMR* (*Differential Microwave Radiometer*), qui mesure des anisotropies d'amplitudes relatives 10^{-5} sur des échelles de l'ordre de 7° (Smoot et al., 1992).

L'étude des anisotropies du CMB, fera depuis, l'objet de nombreuses expériences. En effet elle recèle de nombreuses informations sur l'état de l'univers dans le passé.

1.1.4 Le modèle du Big-Bang

Ces trois observations majeures ont donné le cadre général de la cosmologie moderne. Depuis un état primordial dense et chaud l'univers est en expansion et

son contenu se refroidit : c'est le modèle du Big-Bang. Les autres modèles ont peu à peu tous été abandonnés ne pouvant rivaliser avec la prédictivité du modèle du Big-Bang.

Ce modèle décrit un univers dont le contenu se refroidit avec le temps. On peut distinguer différents changements de phase du contenu en énergie-matière depuis l'origine de l'univers. Je m'intéresse à deux périodes en particulier sur lesquelles je reviendrais plus en détail dans la suite (section 1.2.4 et suivantes). 1) La période d'équivalence, notée t_{eq} , correspond au moment où la densité de matière devient équivalente à la densité de photons. L'évolution de l'univers passe d'une dynamique dominée par l'énergie à une dynamique dominée par la matière. Cette période correspond à 10^4 ans après l'origine du temps. 2) La période de découplage, notée t_{dec} , correspond à l'époque de la recombinaison entre noyaux et électrons pour former les premiers atomes. Le rayonnement et la matière se dissocient, ils se découplent. L'univers est alors transparent aux photons qui peuvent se propager librement.

Le modèle du Big-Bang est décrit dans les ouvrages en anglais de Peebles (Peebles, 1980; Disney, 1993) ou Peacock (Peacock, 1999) et dans le livre plus récent de Rich (Rich, 2002).

1.1.5 Le contenu de l'univers aujourd'hui

Les observations actuelles montrent que la densité totale de l'univers peut être divisée en différentes composantes aux comportements différents : la matière relativiste (rayonnement) ou non (matière) et l'énergie sombre.

Aujourd'hui les mesures des différentes densités montrent que l'univers est constitué de 70% d'énergie sombre, apparentée à une énergie du vide, de 25% de matière sombre, se comportant gravitationnellement comme la matière, et de seulement 5% de matière baryonique, dont moins de 1% directement visible. Je présente dans cette section ces différentes composantes.

1.1.5.1 Les baryons lumineux

Les baryons "lumineux" représentent moins de 1% du contenu en énergie de l'univers. Le principal, et le premier historiquement, moyen de les observer a été l'observation des étoiles en optique. Les observations en X puis grâce à l'effet Sunyaev Zel'dovich (voir le paragraphe 2.4) des amas de galaxies ont permis de constater l'existence d'une grande quantité (en fait beaucoup plus que dans les étoiles) de baryons sous forme de gaz chaud intra-amas. Les baryons visibles, les étoiles des galaxies et le gaz chaud intra-amas, représentent, à peu près, 13% de la masse totale de l'amas. Par ailleurs, on peut mesurer indirectement la quantité de baryons "sombres", décrits dans le paragraphe suivant sous le nom de matière sombre baryonique.

1.1.5.2 La matière sombre

D'un point de vue observationnel, le problème de la matière sombre est lié à la comparaison de la masse réelle d'une structure avec sa masse "apparente". À toutes les échelles (galaxies, amas, cosmologie) il existe trop peu de masse "visible" pour rendre compte des effets dynamiques observés. Par exemple la vitesse de rotation

d'une étoile située en périphérie d'une galaxie, décroît en raison inverse de la racine carrée de la distance au centre de la galaxie. Les observations sont en désaccord avec ces prédictions : on observe des vitesses constantes jusqu'à la périphérie de la galaxie. Zwicky (1933) est le premier à avoir supposé l'existence de la matière sombre en calculant le rapport masse sur luminosité de l'amas de Coma : il mesure à l'époque un rapport : $M/L \simeq 50$ (il a utilisé la valeur de la constante de Hubble de l'époque alors presque dix fois plus grande que celle retenue actuellement). Aujourd'hui ce rapport est mesuré et vaut plutôt 350 (Łokas and Mamon, 2003). Ce rapport M/L est un bon indicateur de la présence de la matière sombre, par convention il vaut 1 pour le soleil : $M_{\odot}/L_{\odot} = 1$. Un rapport plus grand que 1 indique une prédominance de la matière sombre sur la matière "lumineuse", visible. Ostriker et al. (1974) ont montré que la masse manquante pouvait jouer un rôle au niveau cosmologique et que le rapport M/L augmente avec la distance au centre des galaxies.

La nucléosynthèse prédit que la contribution des baryons se situe entre 2% et 5% de la densité critique; les estimations sur la contribution des étoiles (matière lumineuse baryonique) donnent une valeur de 0.5%. Par ailleurs les estimations de la densité de matière (baryonique et sombre) donnent une valeur de 30% de la densité totale. Il apparaît donc que la matière baryonique ne contribue que très peu à la densité de la matière, la matière sombre non-baryonique est largement dominante.

La matière sombre baryonique est invisible car non lumineuse, il faut donc la chercher par d'autres moyens. Une première piste suivie par les expériences MACHO et EROS consiste à rechercher des objets compacts dans la galaxie, dénommés MACHO (Massive Halo Compact Objects). Cette catégorie inclut les naines brunes, les grosses planètes gazeuses, les naines blanches et les étoiles à neutrons. La détection de ces objets se fait par effet de micro-lentille gravitationnelle (Paczynski, 1986; Griest, 1991). Les résultats montrent qu'au plus 50% de la masse du halo serait sous la forme de MACHO (Alcock et al., 1997-1998; Afonso et al., 1999).

Les nuages moléculaires froids et denses sont aussi de bons candidats. Ils ont été détectés pour la première fois par Valentijn and van der Werf (1999). Les mesures ont été effectuées par le satellite ISO et ont prouvé la présence de nuages d'hydrogène moléculaire dans le disque de la galaxie NGC 891. Ce surplus de masse suffirait à résoudre le problème de la masse manquante. Il reste à confirmer ces résultats.

L'existence de la matière sombre non-baryonique, que j'appellerai matière noire, est aujourd'hui bien acceptée par l'ensemble de la communauté. Elle est la forme dominante de la matière et compte pour 25% dans la densité de l'univers. Longtemps les physiciens ont distingué deux types de matière noire : une chaude (HDM pour *Hot Dark Matter*) et une froide (CDM pour *Cold Dark Matter*), les deux formes étant en compétition. La matière noire chaude (relativiste peu après le découplage) impliquerait une formation des structures de type "top-down" (i.e. du haut vers le bas) : les grandes structures se seraient formées en premier pour ensuite se scinder en structures plus petites. L'autre modèle de matière noire froide (CDM, non-relativiste à l'époque de la formation des structures) implique une formation des structures de type "bottom-up" (i.e. du bas vers le haut) : les structures de plus en plus grandes se forment par accréation de structures plus petites. C'est ce dernier modèle qui est retenu aujourd'hui (Peebles and Shaviv, 1982). La distinction s'est faite par la mesure du spectre de puissance des galaxies : on n'observe pas sur ce spectre la chute de puissance aux petites échelles angulaires qu'implique un modèle HDM. Par

ailleurs ce modèle est incapable de décrire la formation des grandes structures sur toute la durée du processus.

Un certain nombre de particules sont retenues comme éventuel candidat pour expliquer la matière noire froide CDM. Elles sont issues des théories de supersymétrie, de grande unification (GUT), des supercordes. Les WIMPS (pour *Weakly Interacting Massive Particles*, issus des théories supersymétriques) en sont les meilleurs candidats. Parmi les WIMPS on peut citer le neutralino, la plus légère des particules supersymétriques de masse 10 GeV - 3 TeV (voir Jungman et al. (1995) pour une revue de la matière noire supersymétrique).

1.1.5.3 L'énergie sombre

L'énergie sombre est apparue dans les équations de la relativité générale sous la forme du terme de la constante cosmologique Λ (voir l'équation : 1.7). A l'origine Einstein l'a introduite comme un effet purement géométrique afin de rendre l'univers stationnaire, mais les observations devenant de plus en plus précises (notamment les observations des supernovae de type Ia), on l'interprète plutôt aujourd'hui comme une énergie du vide (elle a la dimension d'une énergie). L'énergie sombre est assimilée à un fluide de pression négative. L'énergie de cette composante est indépendante d'une particule précise, elle n'est pas diluée par l'expansion cosmologique et reste constante au cours du temps.

1.2 Description de l'univers homogène et isotrope

Dans le cadre du Big-Bang l'univers a été initialement très dense. A cette époque les 4 forces d'interactions ont eu une influence sur sa structure et son contenu. Aujourd'hui seules les forces à longues portées agissent dans l'univers. Les charges mises en jeu à des grandes échelles s'écrantent et la force électromagnétique n'entre pas en jeu pour l'évolution de l'univers. La gravitation est la seule force qui a aujourd'hui (et depuis longtemps!) un rôle dans l'évolution de l'univers. La meilleure théorie de la gravitation est la description géométrique de cette interaction : la relativité générale d'Einstein.

1.2.1 Le principe cosmologique

Le principe cosmologique est l'hypothèse selon laquelle l'univers est homogène et isotrope à grande échelle. Ce principe est vérifié par les expériences de mesure du CMB : les variations sont de l'ordre de 10^{-5} , et par les expériences de sondage profond de galaxies comme le SDSS (*Sloan Digital Sky Survey*) (Tegmark et al., 2004) ou le 2dFGRS (*2 degrees Fields Galaxy Redshift Survey*) (Madgwick et al., 2003) qui montrent aux échelles du *Mpc* des structures filamenteuses mais qui tendent à disparaître aux échelles du *Gpc*.

L'évolution de l'univers en fonction de son contenu en énergie-matière est déduite des équations de la relativité générale : ce sont les équations de Friedmann-Lemaître. L'hypothèse du principe cosmologique impose la forme de la métrique et réduit le nombre de cosmologies possibles.

1.2.2 Relativité et gravitation

En partant du fait la vitesse de la lumière est une grandeur finie Einstein a mis en place la théorie de la relativité restreinte. L'univers est alors décrit comme un espace plat Minkowskien dans lequel les distances entre deux événements (ds) sont données par l'équation :

$$ds^2 = c^2 dt^2 - (dx^2 + dy^2 + dz^2) \quad (1.1)$$

où ds est la distance élémentaire dans cet espace, dx , dy et dz les intervalles d'espace et dt l'intervalle de temps. Pour une ligne de vie de type lumière on a : $ds = 0$. Un objet se déplaçant à une vitesse inférieure à celle de la lumière suit une ligne de vie "temporelle" avec $ds > 0$.

En théorie de la relativité générale l'espace reste localement Minkowskien mais la métrique peut être courbe, c'est un espace Riemanien. La distance élémentaire est alors :

$$ds^2 = g_{\alpha\beta} dx^\alpha dx^\beta \quad (1.2)$$

où $g_{\alpha\beta}$ s'appelle le tenseur métrique et ses composantes dépendent de la gravitation. En relativité générale le contenu en énergie-impulsion de l'espace déforme sa géométrie. L'équation d'Einstein exprime ce fait. Elle s'écrit :

$$R_{\alpha\beta} - \frac{1}{2} g_{\alpha\beta} R = 8\pi G T_{\alpha\beta} \quad (1.3)$$

La partie gauche de l'équation est reliée à la métrique de l'univers (sa forme) et le membre de droite $T_{\alpha\beta}$ est le tenseur d'énergie-impulsion du contenu de l'espace considéré. Les indices α et β décrivent l'ensemble des dimensions : spatiales et temporelle.

La démarche permettant de décrire en relativité générale l'évolution de l'univers est la suivante :

1. les considérations d'isotropie et d'homogénéité de l'univers indiquent la forme de la métrique,
2. l'observation du contenu de l'univers donne la forme du tenseur d'énergie-impulsion,
3. l'équation d'Einstein 1.3 permet de déterminer la dynamique du contenu de l'univers mais aussi de sa forme, les deux étant étroitement liées.

1.2.3 La métrique de Robertson-Walker

Le principe cosmologique fixe la forme de la métrique de l'univers, la métrique de Robertson-Walker. En effet l'isotropie de l'univers (observée) impose à la métrique spatiale une symétrie sphérique. L'équation générale 1.2 peut s'écrire dans ce cas, en coordonnées sphériques, sous la forme :

$$ds^2 = c^2 dt^2 - a^2(t) \left(\frac{dr^2}{1 - kr^2} + r^2(d\theta^2 + \sin^2 \theta d\phi^2) \right) \quad (1.4)$$

où l'on a introduit deux grandeurs importantes car donnant directement une indication sur la forme de l'univers et son évolution :

- $k = -1; 0; 1$ la courbure spatiale (hyperbolique pour un univers ouvert, nulle pour un univers plat ou sphérique pour un univers fermé) de l'univers (considérée constante dans l'espace)
- $a(t)$ le facteur d'échelle donnant la "forme temporelle" (l'évolution) de l'univers.

1.2.4 Le tenseur énergie-impulsion

Le tenseur énergie-impulsion décrit le contenu de l'univers. On se place dans le cadre d'un univers rempli d'un fluide parfait. Le tenseur énergie-impulsion est alors le suivant :

$$T_{\alpha\beta} = (\rho + P)u_\alpha u_\beta - P g_{\alpha\beta} \quad (1.5)$$

où P est la pression propre du fluide parfait, ρ sa densité propre moyenne, u_α est sa quadri-vitesse. La conservation de l'énergie se traduit en relativité générale par la conservation du tenseur d'énergie-impulsion, ce qui donne l'équation suivante (équivalente à celle obtenue en appliquant le premier principe de la thermodynamique à un volume d'univers en expansion) :

$$\dot{\rho} + 3\frac{\dot{a}}{a}(\rho + \frac{P}{c^2}) = 0 \quad (1.6)$$

La densité dans cette équation est la densité d'énergie propre moyenne du contenu de l'univers divisée par c^2 .

Domination	équation d'état	ρ	a	époque
Constante cosmologique	$w = -1$	-	e^{Ht}	inflation
Rayonnement	$w = 1/3$	$\propto 1/a^4$	$t^{1/2}$	avant t_{eq}
Matière	$w = 0$	$\propto 1/a^3$	$t^{2/3}$	après t_{eq}
Courbure	-	-	t	?

TAB. 1.1 – Paramètres de l'équation d'état du contenu de l'univers et les différentes ères de l'univers.

1.2.5 La dynamique de Friedmann-Lemaître

Nous avons décrit le contenu de l'univers et la forme générale de sa métrique. Qu'en est il de sa dynamique ? L'équation d'Einstein 1.3 aboutit aux équations de Friedmann-Lemaître (composantes spatiale et temporelle de l'équation d'Einstein) :

$$\left(\frac{\dot{a}}{a}\right)^2 = \frac{8\pi G}{3}\rho - \frac{kc^2}{a^2} + \frac{\Lambda c^2}{3} \quad (1.7)$$

$$\frac{\ddot{a}}{a} = -\frac{4\pi G}{3}(\rho + 3\frac{P}{c^2}) + \frac{\Lambda c^2}{3}$$

où a est le facteur d'échelle, G la constante gravitationnelle, ρ la densité moyenne de matière-énergie, k la courbure de l'espace, c la vitesse de la lumière dans le vide et Λ la constante cosmologique, où énergie du vide (voir le paragraphe 1.1.5.3).

Ces équations décrivent la dynamique de l'univers et permettent d'obtenir les expressions de $a(t)$, $\rho(t)$ et $P(t)$. En notant que ces deux équations ont pour conséquence l'équation de la conservation de l'énergie 1.6, et que nous avons trois inconnues, il apparaît qu'il manque une équation, l'équation d'état reliant $\rho(t)$ et $P(t)$:

$$P = w\rho c^2 \quad (1.8)$$

où w dépend du type de fluide considéré : matière relativiste, non-relativiste, constante cosmologique. Les variations de densité de ces différentes composantes donnent l'évolution globale de l'univers et permettent de distinguer différentes périodes résumées dans le tableau 1.2.5. Dans ce tableau on introduit différentes époques :

- la période d'équivalence t_{eq} . Cette époque correspond au moment où les densités de matière et de rayonnement sont égales. Avant, le rayonnement domine la dynamique de l'univers, après c'est la matière qui domine.
- la période d'inflation est une période de l'univers primordial où l'expansion a été extrêmement violente (voir le paragraphe 1.3.1).

La période d'équivalence revêt pour nous une grande importance : quelque temps après ce moment, à la période du *découplage* le CMB sera émis et la formation des grandes structures commencera. Ce point constitue le sujet physique principal de ma thèse et est décrit en détail à la fin de ce chapitre et au chapitre 2.

1.2.6 Des quantités importantes pour la suite

J'introduis ici des quantités importantes pour la compréhension des chapitres suivants sur la formation des amas de galaxies. Ces quantités permettent de faire le

lien entre les équations de la relativité générale et les mesures que peut la cosmologie observationnelle.

Distance propre Afin de bien comprendre ce qu'est la constante (non constante !) de Hubble on introduit la notion de distance propre. Considérons deux points de l'espace au même instant t . En reprenant la métrique de Robertson-Walker la distance propre entre ces deux points est :

$$d_p(t) = \sqrt{|ds^2|} = a(t) \int_0^r \frac{dr'}{\sqrt{1 - kr'^2}} = a(t)f(r) \quad (1.9)$$

où $f(r) = r$ si $k = 0$, $f(r) = \sin^{-1} r$ si $k = 1$ et $f(r) = \sinh^{-1} r$ si $k = -1$. La grandeur $f(r)$ est la distance comobile entre les deux points. Dans un univers en expansion cette grandeur reste constante mais la distance propre suit (et donc mesure) l'expansion de l'univers. La vitesse entre ces deux points s'exprime de la manière suivante :

$$v(t) = \dot{a}(t)g(r) = \frac{\dot{a}}{a}d_p(t) \equiv H(t)d_p(t) \quad (1.10)$$

où $H = \frac{\dot{a}}{a}$ est la constante de Hubble. Sa valeur actuelle $H_0 = H(t = t_0)$ est donnée sous forme réduite par :

$$h = \frac{H_0}{100 \text{ km/s/Mpc}} = 0.72 \pm 0.08 \quad (1.11)$$

Redshift Dans le paragraphe précédent on remarque que la notion de distance pose un problème en cosmologie : elle dépend justement du cadre cosmologique dans lequel on se place. On utilise plutôt, pour mesurer les distances, le *redshift*. C'est une grandeur directement mesurable (bien que ça ne soit pas forcément simple) des objets astrophysiques. Quelque soit le cadre cosmologique que l'on considère le *redshift* mesuré d'une galaxie est défini de la manière suivante :

$$z = \frac{\lambda_0 - \lambda_e}{\lambda_0} \quad (1.12)$$

où λ_e est la longueur d'onde d'un signal lumineux envoyé par un émetteur éloigné. λ_0 est la longueur d'onde que l'observateur mesure de ce même signal. On constate que le paramètre d'expansion est relié au *redshift* par la relation :

$$a(z) = \frac{a_0}{(1 + z)} \quad (1.13)$$

Dans la suite on utilisera le *redshift* comme indicateur de temps et on écrira indifféremment $a(z)$ ou $a(t)$, $H(z)$ ou $H(t)$, etc.

Les paramètres cosmologiques On introduit les paramètres cosmologiques usuels en réécrivant les équations de Friedmann-Lemaître avec $a(z) = a_0/1 + z$ et $H(z) = \dot{a}(z)/a(z)$:

$$1 = \Omega_M(z) - \Omega_k(z) + \Omega_\Lambda(z) \quad (1.14)$$

où

- $\Omega_M(z) = \frac{8\pi G \rho(z)}{3H^2(z)} = \frac{\rho}{\rho_c}$ est la densité d'énergie du fluide considéré. $\rho_c \equiv 3H^2/8\pi G$ est la densité critique de l'univers.
- $\Omega_k(z) = \frac{kc^2}{a^2(z)H^2(z)}$ est la "densité de courbure" de l'univers.
- $\Omega_\Lambda(z) = \frac{\Lambda c^2}{3H^2(z)}$ est la "densité du vide" (voir la discussion sur le paramètre Λ et l'énergie sombre).

L'équation 1.14 est valable à tous les *redshift*, et notamment à $z = 0$ où l'on note :

$$\Omega_M = \Omega_M(0); \Omega_\Lambda = \Omega_\Lambda(0); \Omega_k = \Omega_k(0) \quad (1.15)$$

ces grandeurs sont ce que l'on appelle les paramètres cosmologiques. Les observations des anisotropies primaires du CMB (Spergel et al., 2003), des supernovae de type Ia (Lampeitl et al., 2004) et les sondages de galaxies (Madgwick et al., 2003) ont permis de contraindre les valeurs de ces paramètres (les résultats sont donnés dans le tableau 1.2.6). On peut exprimer les paramètres Ω_M , Ω_Λ et Ω_k en fonction du *redshift* en introduisant $E(z) = H(z)/H_0$ dans l'équation 1.14, on a alors :

$$1 = \Omega_M(z) - \Omega_k(z) + \Omega_\Lambda(z) \quad (1.16)$$

avec

$$\begin{aligned} \Omega_M(z) &= \Omega_M^{nr}(z) + \Omega_M^r(z) = (\Omega_M^{nr}(1+z)^3 + \Omega_M^r(1+z)^4) E^{-2}(z) \\ \Omega_k(z) &= \Omega_k(1+z)^2 E^{-2}(z) \\ \Omega_\Lambda(z) &= \Omega_\Lambda E^{-2}(z) \end{aligned} \quad (1.17)$$

où les exposants nr et r indiquent respectivement la densité de matière non-relativiste et relativiste. L'équation 1.14 s'écrit alors sous la forme :

$$\left(\frac{H(z)}{H_0} \right)^2 = \Omega_M^{nr}(1+z)^3 + \Omega_M^r(1+z)^4 - \Omega_k(1+z)^2 + \Omega_\Lambda \quad (1.18)$$

Le but de la cosmologie observationnelle est de mesurer les paramètres cosmologiques. Le tableau 1.2.6 donne un état des lieux de la valeur des paramètres cosmologiques aujourd'hui.

Distance angulaire La distance angulaire est par définition le rapport de la distance physique entre deux points d situés à un *redshift* z sur leur distance apparente θ :

$$D_a = \frac{d}{\theta} = a(t)r \quad (1.19)$$

Paramètre	Valeur centrale	Erreur statistique 1σ
Ω_M	0.29	0.04
Ω_b	0.049	0.003
h	0.69	0.03
n_s	0.97	0.03

TAB. 1.2 – Ces valeurs sont tirées de Yeche et al. (2005) pour un univers Λ CDM où $\Omega_{total} = 1$. Elles correspondent aux contraintes calculées grâce à la mesure du spectre de puissance du CMB par WMAP (Spergel et al., 2003) et à l'observation des supernovae de type Ia (Lampeitl et al., 2004).

où r est la distance entre l'observateur et les deux objets. On peut montrer que, quel que soit le modèle cosmologique avec $k = -1; 1$:

$$D_a(z) = \frac{c}{\sqrt{|\Omega_k|}H_0(1+z)} S\left(\sqrt{|\Omega_k|} \int_1^{1+z} \frac{du}{E(z)}\right) \quad (1.20)$$

avec S une fonction dépendante de la valeur de k : pour $k = -1$, S est la fonction sinus hyperbolique et pour $k = 1$, S est la fonction sinus. Pour un univers plat ($k = 0$) on a :

$$D_a(z) = \frac{c}{H_0(1+z)} \int_1^{1+z} \frac{du}{\Omega_\Lambda + \Omega_m u^3} \quad (1.21)$$

Horizon et rayon de Hubble L'horizon est simplement l'expression, en terme de distance, du fait que la lumière a une vitesse finie. Une particule se déplaçant à la vitesse de la lumière va donc parcourir en un temps fini une distance finie : la taille de l'horizon. Nous ne pouvons pas voir les régions se trouvant à une distance supérieure au produit de la vitesse de la lumière par l'âge de l'univers.

La loi de Hubble montre que plus un objet comobile est lointain plus il s'éloigne vite par rapport à un observateur donné. La sphère de Hubble est la distance atteinte par un objet se déplaçant à la vitesse de la lumière. La sphère de Hubble est appelée parfois *horizon cosmologique effectif* car c'est la grandeur pertinente dans les processus physiques pouvant se produire à une échelle donnée ; si une structure a une dimension caractéristique supérieure au rayon de Hubble, seuls les effets d'expansion de l'univers la feront évoluer à travers la métrique. Si sa taille est inférieure, ce sont les processus causaux de la physique qui la feront évoluer. C'est la grandeur de référence utilisée pour décrire l'évolution des surdensités : nous reviendrons sur ce point dans la section 1.3.2.

1.2.7 Une brève histoire thermique

Nous avons vu dans la partie précédente que l'univers est en expansion. À son origine il était extrêmement dense et chaud. Aujourd'hui on observe le CMB à la température extrêmement froide de 3 K. Nous avons ensuite introduit l'époque de

l'équivalence t_{eq} , du découplage t_{dec} (la période d'émission du CMB) etc. Revenons sur cette évolution d'un point de vue thermique. L'énergie d'un photon est donnée par :

$$E = h\nu = \frac{h\lambda}{c} \quad (1.22)$$

Dans un univers en expansion on a :

$$E \frac{1}{a} \quad (1.23)$$

Ainsi, l'énergie des photons, mesurée par un observateur au repos, dans un univers en expansion décroît : c'est le décalage vers le rouge ou *redshift* cosmologique introduit précédemment. Le spectre reste un spectre de corps noir au cours de l'expansion mais sa température diminue :

$$T = T_0(1 + z) \quad (1.24)$$

où T_0 est la température actuelle : 2.725 K. A l'époque du découplage, les photons étaient thermalisés et leur distribution en énergie suit aussi un spectre de corps noir de température :

$$T_{dec} = T_0(1 + z_{dec}) \simeq 3000 \text{ K} \quad (1.25)$$

En procédant de la sorte on peut construire une "thermo-chronologie" de l'univers. Il convient de préciser que depuis l'origine du temps jusqu'à un temps de 10^{-12} s on se trouve en *terra incognita* de la physique. Des théories existent mais elles ne sont pas vérifiables expérimentalement de manière directe.

$t \simeq 10^{-43}$ s **ou** $E \simeq 10^{19}$ GeV. C'est le temps de Planck. Les théories de grande unification des forces fondamentales (GUT) prédisent qu'à cette échelle les quatre interactions (gravitationnelle, électromagnétique, nucléaire faible et nucléaire forte) sont unifiées, mais aucune théorie actuellement ne peut décrire la physique de ces premiers instants de façon satisfaisante. Peu après le temps de Planck se déroule une période d'expansion accélérée de l'univers : l'inflation (voir la section suivante).

$t \simeq 10^{-35}$ s, $T \simeq 10^{26}$ K **ou** $E \simeq 10^{15}$ GeV. A cette période les deux interactions : forte et électrofaible sont unifiées. Aux énergies plus faibles la symétrie se "brise" (mécanisme de Higgs) et les particules sont décrites dans le modèle standard (des particules) par le groupe $SU(3) \times (SU(2) \times U(1))$.

$t \simeq 10^{-12}$ s, $T \simeq 10^{16}$ K **ou** $E \simeq 300$ GeV. A cette énergie se produit la transition de phase électrofaible : la symétrie électrofaible ($SU(2) \times U(1)$) est brisée (mécanisme de Higgs) ce qui différencie les deux interactions.

$t \simeq 10^{-6}$ s, $T \simeq 10^{13}$ K **ou** $E \simeq 100$ MeV. A cette énergie se déroule la transition de phase quark-hadrons. La température, trop basse, ne permet pas aux quarks de rester libres : ils se confinent en hadrons (nucléons et plus généralement particules sensibles à l'interaction forte). Aux températures supérieures l'univers est rempli d'un

plasma de quark-gluons, objet de nombreuses recherches dans les accélérateurs de particules actuelles qui peuvent atteindre ces énergies lors de collisions entre ions lourds (RHIC, LHC, etc).

$t \simeq 1 \text{ s}$ **ou** $E \simeq 1 \text{ MeV}$. A cette température les neutrinos se découplent de la matière : on s'attend à l'émission d'un fond diffus de neutrinos. Peu après les paires de particules matière-antimatière s'annihilent. Il ne reste à la fin qu'une fraction de matière, celle encore présente aujourd'hui. Une question se pose : pourquoi existait-il une différence de densité (très faible) entre ces deux populations (matière et antimatière) à cette période ?

$t \simeq 100 \text{ s}$, $z \simeq 1150$, $T \simeq 3500 \text{ K}$ **ou** $E \simeq 0.1 \text{ MeV}$. Les premiers noyaux de deutérium peuvent se former quand la température est devenue suffisamment basse pour qu'ils ne soient plus photo-dissociés. C'est la nucléosynthèse primordiale, dont le paramètre principal est le rapport η du nombre de baryons au nombre de photons. Ce paramètre est constant au cours du temps et vaut environ 10^{-10} . Les éléments plus lourds se forment ensuite : tritium, hélium 3 et 4, puis lithium 6 et 7. Les abondances de ces différents éléments peuvent être calculées et elles sont en très bon accord avec les observations.

$t \simeq 10^4 \text{ ans}$, $z \simeq 1150$, $T \simeq 3500 \text{ K}$ **ou** $E \simeq 0.5 \text{ eV}$. C'est l'époque de l'équivalence rayonnement-matière. L'univers passe d'une ère dont la dynamique est dominée par le rayonnement vers une ère où la matière domine : le facteur d'échelle passe du régime $a \propto t^{1/2}$ au régime $a \propto t^{2/3}$.

$t \simeq 5 \times 10^5 \text{ ans}$, $z \simeq 1000$, $T \simeq 3000 \text{ K}$ **ou** $E \simeq 0.3 \text{ eV}$. C'est l'époque du découplage. Les électrons se recombinent avec les noyaux pour former les premiers atomes : rayonnement et matière se dissocient, l'univers devient transparent aux photons (leur libre parcours moyen dépasse la taille de l'univers). Il est intéressant de noter que l'énergie à l'époque du découplage n'est pas celle d'ionisation de l'atome d'hydrogène 13.6 eV . Ceci est dû au fait que le spectre de corps noir est très étalé et que les photons sont largement en surnombre par rapport aux baryons. Il faut attendre une énergie de 0.3 eV pour que le milieu cosmologique cesse d'être ionisé. C'est la période d'émission du CMB. C'est ce rayonnement que nous observons aujourd'hui à la température beaucoup plus basse de 2.725 K (nous reviendrons sur ce point en fin de chapitre).

$t \simeq 10^9 \text{ ans}$, $z \simeq 5$ **ou** $T \simeq 15 \text{ K}$. La matière domine l'univers et la gravitation est l'interaction prédominante. Les grandes structures se forment suivant un scénario "bottom-up". Les amas de galaxies apparaissent par accrétion de gaz et de structures de petite taille.

$t \simeq 13.5 \times 10^9 \text{ ans}$, $z = 0$ **ou** $T \simeq 3 \text{ K}$, **aujourd'hui**. L'évolution de l'univers est gouvernée par la gravitation et par l'énergie noire. L'univers est isotrope et homogène aux échelles supérieures au Gpc . La gravitation a créé des contrastes de den-

sité extrêmement forts entre les régions sur-denses et les régions sous-denses. Nous sommes plongés dans un rayonnement de corps noir, le CMB, à une température de 2.725 K.

1.3 L'univers inhomogène : de l'inflation au découplage

Le principe cosmologique postule l'isotropie de l'univers aux grandes échelles. C'est ce que l'on observe grâce aux observations du CMB et aux expériences d'observation de galaxies du type 2dF ou SDSS. En dessous de cette échelle l'univers nous apparaît fortement inhomogène : présence d'amas de galaxies, de filaments, de fluctuations de température dans le CMB, etc. Dans cette section je suis un développement chronologique : en partant de l'inflation, modèle prédisant l'apparition de fluctuations de densité cohérentes dès l'univers primordial, je suivrai l'évolution du champ de densité jusqu'à la période du découplage, pour aboutir à la description des deux composantes (découplées) de l'univers que l'on observe aujourd'hui : le rayonnement, et la matière, notamment les grandes structures massives que sont les amas, sujet du chapitre suivant.

1.3.1 L'inflation

Nous nous intéressons ici à l'origine des fluctuations de densité primordiales (elles-mêmes à l'origine de celles que l'on observe aujourd'hui). Le principal apport du modèle de l'inflation (voir pour plus de détails Rich (2002) page 169), phase primordiale d'expansion accélérée de l'univers, est la production de fluctuations de densité spatialement cohérentes obéissant à une statistique gaussienne (la recherche de non-gaussianité primordiale dans les fluctuations du CMB est un domaine en plein essor de la cosmologie observationnelle : la difficulté provient de la présence d'un bruit d'avant plan de forte amplitude : SZ, weak-lensing, etc.). Le modèle de l'inflation a été proposé au début des années 80 (Guth, 1981) pour répondre à trois problèmes du modèle standard de la cosmologie de l'époque découlant des observations.

La platitude. Le premier problème rencontré est la platitude de l'univers. Les données de WMAP ont montré avec une grande précision que l'univers est quasiment plat aujourd'hui : $\Omega_M + \Omega_\Lambda \sim 1$ soit $\Omega_k \sim 0$. En reprenant les équations de Friedman-Lemaître on voit qu'il est alors nécessaire d'avoir un univers *initialement* presque parfaitement plat. On se pose alors la question : pourquoi un univers initialement plat ?

L'horizon. Le second problème soulevé est celui de l'horizon. L'horizon, comme nous l'avons vu, est défini comme la distance maximale de laquelle un observateur peut recevoir de l'information. Au moment de l'émission du CMB on trouve un horizon ayant une taille apparente de un degré. Comment le CMB peut-il présenter des températures homogènes (à 10^{-5} près) sur l'ensemble du ciel alors que les régions liées causalement avant son émission sont espacées au maximum de un degré ?

L'origine des fluctuations de densité. Il apparaît aujourd'hui que l'amplitude des fluctuations observées est incompatible avec une évolution standard de l'univers, dominée par le rayonnement puis par la matière. On montre, dans la section suivante, que les fluctuations croissent uniquement à l'intérieur du rayon de Hubble : dans

ce cas les grandes structures n'auraient pas eu le temps de se former. Elles ont donc commencé à se former avant, à l'extérieur du rayon de Hubble. Comment les fluctuations ont-elles pu croître avant le découplage rayonnement-matière ?

Il est possible de répondre à ces trois questions en considérant que l'expansion de l'univers s'est faite de manière accélérée pendant un temps très bref, juste après le big-bang. On aurait alors le scénario suivant :

1. L'expansion de l'univers est dominée par le rayonnement jusqu'à l'entrée dans la période de l'inflation : de $t = 0$ à $t = t_i$,
2. puis commence une période de croissance exponentielle jusqu'à la fin de l'inflation $t = t_f$,
3. ensuite l'univers est à nouveau dominé par le rayonnement, jusqu'à l'équivalence $t = t_{eq}$,
4. enfin, après t_{eq} , l'univers est dominé par la matière.

Le problème de la platitude est naturellement résolu dans ce modèle. En effet pendant l'inflation on a $\ddot{a} > 0$ (expansion accélérée), soit \dot{a} augmente et donc $Ha = \dot{a}$ augmente avec le temps, ce qui se traduit par l'évolution :

$$|\Omega_M + \Omega_\Lambda - 1| = \Omega_k = \frac{|k|c^2}{H^2 a^2} \rightarrow 0 \quad (1.26)$$

Un univers où la constante cosmologique domine tend asymptotiquement vers un univers plat. La période d'inflation, pendant laquelle la dynamique de l'évolution est dominée par la constante cosmologique, fait tendre la géométrie de l'univers vers celle d'un univers plat.

Le problème de l'horizon devient caduque dans le contexte de l'inflation : deux points du CMB éloignés de plus d'un degré étaient, avant l'inflation, causalement liés et ont donc pu échanger de l'information et homogénéiser leur température.

Par ailleurs l'inflation permet de générer un champ stochastique gaussien de fluctuations de densité. C'est ce champ qui est à l'origine des grandes structures que nous observons aujourd'hui.

Bien que le modèle d'inflation nécessite l'introduction de champs scalaires *ad-hoc* non expliqués aujourd'hui, il est généralement admis par la communauté scientifique, restant toutefois l'objet de tests expérimentaux.

1.3.2 Évolution du champ de densité

Nous avons vu que l'inflation génère de façon naturelle les fluctuations gaussiennes du champ de densité (que l'on considérerait jusqu'à présent uniforme). Dans cette partie nous étudions l'évolution des fluctuations de densité sous l'effet de la gravitation dans le cadre de l'univers en expansion.

On se place pour le moment dans les approximations suivantes :

- Nous décrirons ici l'évolution des fluctuations grâce à la mécanique Newtonnienne. Nous considérons des champs gravitationnels faibles, des perturbations dont la dimension caractéristique est plus petite que l'horizon, et des fluides

non-relativistes. Ces approximations sont tout à fait valables pour décrire l'univers dans sa période de transition entre rayonnement et matière : les fluctuations peuvent être considérées comme des perturbations (régime linéaire) et le rayonnement ne domine plus ; on est dans le cas de fluide non-relativiste.

– Matière et rayonnement forment un fluide parfait (sans dissipation).

On cherche ici à décrire l'évolution du contenu de l'univers entre la fin de la période d'inflation jusqu'à la période du découplage rayonnement-matière.

Les questions qui se posent alors sont :

1. Comment se sont formées ces fluctuations ?
2. Comment évoluent elles ?

Les fluctuations Les fluctuations sont décrites par une perturbation de la densité moyenne, de la pression moyenne et du champ gravitationnel moyen :

$$\begin{aligned}\rho(t, r) &= \bar{\rho}(t) + \delta\rho(t, r) \\ p(t, r) &= \bar{p}(t) + \delta p(t, r) \\ \phi(t, r) &= \bar{\phi}(t) + \delta\phi(t, r)\end{aligned}\tag{1.27}$$

En effet à la période du découplage rayonnement-matière le CMB est émis et les perturbations de température mesurées (et donc de densité) sont de l'ordre de 10^{-5} . Avant et peu après cette période on suppose donc $\delta\rho \ll 1$, $\delta p \ll 1$, $\delta\phi \ll 1$. Pour des raisons de commodité on pose dans la suite : $\delta = \delta\rho$.

Les équations d'évolution pour un fluide parfait dans un champ de gravité sont simplement les équations de Euler et de conservation de la masse.

Il est important de noter qu'en pratique on traitera de deux types de fluide : celui composé de matière noire, sans collision (sans pression) et qui interagit uniquement via la gravitation. Le deuxième composé de la matière baryonique interagit avec le premier par gravitation et avec les photons par diffusion de Compton (pression non nulle).

1.3.2.1 La matière noire, un fluide à pression nulle

On considère dans ce qui suit un fluide de matière sombre de densité ρ_m sans pression. Les seules forces agissantes sont gravitationnelles et attractives.

Étant donné l'expansion de Hubble, on se place dans les unités comobiles de longueur et de vitesse propre :

$$\begin{cases} x = ar \\ \dot{x} = Hx + au \end{cases}\tag{1.28}$$

où r est la coordonnée comobile et $u = \dot{r}$ la vitesse propre comobile. Le point représente la dérivée totale $\frac{d}{dt} \neq \frac{\partial}{\partial t}$. La vitesse d'une particule est la somme du *flot* de Hubble Hx et de sa vitesse propre au .

L'équation d'évolution s'obtient en décomposant l'équation d'Euler sous la forme :

$$\ddot{x} = g_0 + g \quad (1.29)$$

où g est l'accélération propre et g_0 est l'accélération qui agit sur une particule dans un univers de densité uniforme. Par ailleurs l'expression de l'accélération d'une particule se déduit directement en dérivant sa vitesse (1.28) :

$$\ddot{x} = a\dot{u} + 2\dot{a}u + \frac{\ddot{a}}{a}x = g_0 + g \quad (1.30)$$

L'équation sans perturbation ($u = \dot{u} = 0$), est : $\frac{\ddot{a}}{a}x = g_0$; l'équation d'Euler pour les perturbations s'écrit donc simplement :

$$\dot{u} + 2\frac{\dot{a}}{a}u = \frac{1}{a}g \quad (1.31)$$

On peut par ailleurs relier la vitesse d'une particule fluide à la densité par l'équation de conservation de la masse :

$$\frac{d}{dt}\rho + \rho\nabla \cdot u = 0 \quad (1.32)$$

En considérant l'expression de la densité donnée dans l'équation 1.27 et en se plaçant dans un régime quasi-stationnaire (dans lequel l'évolution moyenne est supposée constante sur un cours laps de temps : $\bar{\rho}_m$ indépendant du temps) on obtient :

$$\frac{d}{dt}\delta + (1 + \delta)\nabla \cdot u = 0 \quad (1.33)$$

Cette équation non-linéaire décrit simplement la conservation de la masse en unités comobiles.

A ce stade de l'étude il est nécessaire de linéariser les équations. On considère que les perturbations de densité sont faibles : on se place dans le régime où le rayonnement domine ou peu après (typiquement jusqu'au découplage). Après le découplage, des structures virialisées se forment et le système ne peut plus être décrit par des équations linéaires. On suppose $\delta \ll 1$ et on néglige les termes de second ordre ($\delta \times u$). Cette approximation permet d'écrire :

$$\frac{d}{dt} = \frac{\partial}{\partial t} + u \cdot \nabla \simeq \frac{\partial}{\partial t} \quad (1.34)$$

Finalement les perturbations de densité sont décrites entièrement par le système d'équations linéaires couplées suivant :

$$\begin{cases} \dot{u} + 2\frac{\dot{a}}{a}u = \frac{1}{a}g \\ \dot{\delta} = -\nabla \cdot u \end{cases} \quad (1.35)$$

L'élimination de u dans les équations conduit à l'équation d'évolution des perturbations de densité. On obtient ce résultat en prenant la divergence de l'équation d'Euler et la dérivée par rapport au temps de l'équation de conservation de la masse.

$$\ddot{\delta} + 2\frac{\dot{a}}{a}\dot{\delta} - \frac{1}{a}\nabla g = 0 \quad (1.36)$$

Dans le cadre de la mécanique Newtonienne et en ne considérant que des forces gravitationnelles on a :

$$g = -\frac{\nabla\Phi}{a} \quad (1.37)$$

En utilisant l'équation de Poisson (Newtonienne) $\nabla^2\Phi = 4\pi G\rho$, on obtient :

$$\ddot{\delta} + 2\frac{\dot{a}}{a}\dot{\delta} - 4\pi G\bar{\rho}_m\delta = 0 \quad (1.38)$$

On remarque que le terme en $\dot{\delta}$ est l'équivalent d'un frottement freinant la croissance des surdensités. Il est produit par l'expansion de l'univers paramétrée par $\frac{\dot{a}}{a} = H(t)$.

Regardons le comportement de ce fluide aux différentes périodes de l'univers :

1. Domination de la matière : $\Omega_M^r = 1$.

Dans ce cas on calcule les solutions suivantes :

$$\begin{cases} a(t) \propto t^{2/3} \\ \delta_+ \propto t^{2/3} \propto a \\ \delta_- \propto t^{-1} \propto a^{-3/2} \end{cases} \quad (1.39)$$

Les indices \pm correspondent aux solutions indépendantes croissantes et décroissantes. Avec le temps les perturbations vont s'accroître proportionnellement au facteur d'échelle a . Les modes décroissants disparaissent rapidement.

2. Domination du rayonnement : $\Omega_M^r = 1$.

Dans ce cas, la perturbation δ de la matière non relativiste est négligeable et on obtient les solutions suivantes :

$$\begin{cases} a(t) \propto t^{1/2} \\ \delta_1 = \text{constante} \\ \delta_2 \propto -\ln t \end{cases} \quad (1.40)$$

avec δ_1 et δ_2 différentes solutions possibles. Dans une région dominée par le rayonnement on n'observe pas de modes croissants. L'expansion se fait de manière trop rapide : les surdensités sont diluées plus rapidement qu'elles ne se condensent.

3. Domination du "vide" : $\Omega_\Lambda = 1$ Dans ce cas on a les solutions suivantes :

$$\begin{cases} a(t) \propto e^{Ht} \\ \delta_1 = \text{constante} \\ \delta_2 \propto a^{-2} \end{cases} \quad (1.41)$$

Il n'y a pas de modes croissants : la dilution par l'expansion est trop importante.

Ces résultats sont valables dans le cas d'un fluide à pression nulle comme la matière noire. Regardons ce qui se passe dans le cas du fluide baryons-photons à pression non-nulle.

1.3.2.2 Description du mélange baryons-photons

On considère maintenant la composante baryons-photons. Elle peut-être modélisée par le fluide précédent en incluant une force de pression due à la diffusion Compton des baryons sur les photons. L'accélération d'une particule de ce fluide sera alors $g = -\nabla\Phi/a - \nabla P/(a\bar{\rho})$. Intuitivement le comportement d'un tel fluide sera principalement décrit par une échelle caractéristique à laquelle s'équilibrent pression et gravité. En dessous de cette échelle la pression "résiste" à la gravité, au delà, la gravité domine et les surdensités croissent. Cette longueur est la longueur de Jeans. L'équation générale du mouvement dans ce cas est la suivante (voir l'équation 1.36) :

$$\ddot{\delta} + 2\frac{\dot{a}}{a}\dot{\delta} - \frac{1}{a}\nabla \cdot (-\nabla\Phi/a - \nabla P/(a\bar{\rho})) = 0 \quad (1.42)$$

Les variations de pression sont directement reliées aux variations de densité par la vitesse du son dans le milieu définie de la manière suivante :

$$c_s^2 \equiv \frac{\partial P}{\partial \rho} \quad (1.43)$$

On a alors :

$$p = \bar{p} + \delta p = \bar{p} + \delta \rho \left(\frac{\delta p}{\delta \rho} \right) = \bar{p} + c_s^2 \bar{\rho} \delta \quad (1.44)$$

Par ailleurs les perturbations de densité peuvent être décrites, dans un espace plat, par une superposition d'ondes planes de densité :

$$\delta \equiv \delta(t) e^{-ik \cdot r} \quad (1.45)$$

où la dépendance temporelle est portée par l'amplitude. Le vecteur d'onde k est en unité comobile : la longueur d'onde d'un mode grandit avec l'expansion de Hubble. En utilisant les expressions (1.27) et (1.43) pour exprimer explicitement la pression, on obtient la relation suivante :

$$\ddot{\delta} + 2\frac{\dot{a}}{a}\dot{\delta} + (-4\pi G\bar{\rho} + c_s^2 k^2/a^2) \delta = 0 \quad (1.46)$$

Cette équation a le mérite de faire apparaître très clairement la compétition entre le terme de gravitation $4\pi G\bar{\rho}(t)$ et le terme de pression acoustique $k^2 c_s^2/a^2$. La gravitation intervient comme source avec un signe positif et tend à effondrer la matière, contrairement à la pression qui tend à la disperser. Comme on pouvait s'y attendre il existe une longueur caractéristique pour laquelle les deux contributions sont égales : c'est l'échelle de *Jeans*. Elle est caractérisée par le nombre d'ondes k_J défini par :

$$4\pi G\bar{\rho} = \frac{k_J^2 c_s^2}{a^2} \quad (1.47)$$

soit, en terme de taille spatiale :

$$a\lambda_J = c_s \sqrt{\frac{\pi}{G\bar{\rho}}} \quad (1.48)$$

Une sphère de cette taille contient une masse moyenne M_J :

$$M_J = \frac{4}{3}\pi(a\lambda_J)^3\bar{\rho} \quad (1.49)$$

En faisant l'hypothèse d'un système quasi-stationnaire (c'est-à-dire que a et $\bar{\rho}$ varient peu dans le temps par rapport aux perturbations), cette équation se résout simplement :

- pour $k > k_J$: le coefficient du terme constant est positif et $-4\pi G\bar{\rho}_m + c_s^2 k^2/a^2 \simeq c_s^2 k^2/a^2$ le système est décrit par une équation différentielle du second ordre à coefficients constants :

$$\ddot{\delta} + 2H\dot{\delta} + \frac{c_s^2 k^2}{a^2} = 0 \quad (1.50)$$

Dans ce cas, une solution générale est :

$$\delta = \left(A \cos\left(\frac{c_s k}{a}t\right) + B \sin\left(\frac{c_s k}{a}t\right) \right) e^{-Ht} \quad (1.51)$$

La force de gravité est au plus comparable à la force de pression : on observe des oscillations amorties par l'expansion de l'univers.

- pour $k < k_J$: le coefficient du terme constant est négatif et $-4\pi G\bar{\rho}_m + c_s^2 k^2/a^2 \simeq -4\pi G\bar{\rho}_m$. En supposant $\Omega_m \simeq 1$, le système est décrit par l'équation différentielle du second ordre à coefficients constants :

$$\ddot{\delta} + 2H\dot{\delta} - \frac{3}{2}H^2 = 0 \quad (1.52)$$

dont la solution générale est :

$$\delta = A e^{H(\sqrt{2}-1)t} + B e^{-H(1+\sqrt{2})t} \quad (1.53)$$

Les surdensités croissent.

On observe ainsi deux régimes possibles : le premier pour les longueurs d'onde k supérieures à la longueur d'onde de Jeans $k_J = \sqrt{\frac{4\pi G\bar{\rho}a^2}{c_s^2}}$ (i.e. les échelles $l < l_J = \frac{2\pi a}{k_J} = c_s \sqrt{\frac{\pi}{G\bar{\rho}}}$) où la pression est effective et empêche l'effondrement de la matière. Le second pour les échelles $l > l_J$, où la pression n'est plus prépondérante et l'effondrement gravitationnel des perturbations est possible : les surdensités croissent. Il est important de noter que la longueur de Jeans varie avec le temps et dépend du contenu "moyen" de l'univers ($\bar{\rho}$).

On met en évidence un phénomène important lors de l'évolution des fluctuations : ce sont les oscillations acoustiques. Il est possible de voir directement ces oscillations à travers le CMB, ce sont les pics présents dans le spectre de puissance. Les perturbations de densité se traduisent sur les photons par une perte ou un gain d'énergie (selon qu'ils sortent ou tombent dans un puits de potentiel), tandis que la vitesse du fluide induit un effet Doppler sur les photons.

1.3.2.3 Echelle caractéristique de l'évolution des surdensités

Il apparaît donc dans les calculs que les perturbations de densité du fluide de matière noire ne peuvent s'amplifier que dans une période de domination de la matière. De plus pendant cette période le fluide baryon-photon verra croître ses régions dont les tailles sont plus grandes que l'échelle de Jeans mais seulement après la période de découplage entre ses deux composantes baryons et photons.

D'après ce qui précède on voit que le comportement des perturbations de densité (croissance ou "gel") est directement lié à la longueur de Jeans. Comment celle-ci varie dans le temps et plus particulièrement, avant et après la période d'équivalence rayonnement-matière ? Par définition :

$$l_J(t) \equiv \frac{2\pi a}{k_J} = c_s(t) \sqrt{\frac{\pi}{G\bar{\rho}(t)}} \quad (1.54)$$

où la densité considérée est celle moyenne de l'univers : matière et rayonnement. En considérant l'expression de Ω_{m+r} :

$$\Omega_{m+r} = \frac{8\pi G\bar{\rho}(t)}{3H^2(t)} \simeq 1 \quad (1.55)$$

on en déduit l'expression dépendante du temps de la longueur de Jeans :

$$l_J(t) \simeq \sqrt{\frac{8}{3}} \pi \frac{c_s}{c} d_H(t) \quad (1.56)$$

- *Avant l'équivalence.* Pendant cette période la densité moyenne de l'univers est proche de celle du rayonnement $\bar{\rho} \simeq \bar{\rho}_r$. La vitesse du son vaut $c_s = c/\sqrt{3}$. La longueur de Jeans est alors :

$$l_J(t) \simeq \frac{2\sqrt{2}\pi}{3} d_H(t) \propto a^2 \quad (1.57)$$

La croissance des perturbations ne peut se faire à l'intérieur de l'échelle de Jeans, c'est à dire, dans ce cas à l'intérieur de l'horizon. L'interprétation Newtonienne, dans laquelle la gravitation s'exerce instantanément, indique que les perturbations aux échelles plus grandes peuvent croître. Cependant en considérant le point de vue de la relativité générale, les échelles plus grandes que l'horizon ne peuvent échanger d'information et ces perturbations ne peuvent croître. Pendant cette période les perturbations ne s'amplifient pas : ni celles du fluide de matière noire, ni celles du fluide baryon-photon.

- *Entre l'équivalence et le découplage.* Après l'équivalence la matière domine et l'expansion suit la loi $a \propto t^{2/3}$, les perturbations de matière noire peuvent commencer à croître. Les surdensités de matière noire se creusent et atteignent un contraste de l'ordre de 10^{-3} au moment du découplage.

En ce qui concerne le fluide baryon-photon l'évolution va dépendre du rapport local entre les densités de baryons et de photons. Si la densité de baryon est importante on aura $c_s \propto \rho_r/\bar{\rho} \propto a^{-1}$ et la longueur de Jeans devient $l_J(t) \propto a$,

c'est à dire constante en coordonnées comobiles. Les perturbations décrites par $d_H > l > l_J$ pourront croître. Si la densité de baryon est faible par rapport à celle de photon on aura toujours $l_J(t) \propto d_H(t)$ et la perturbation ne pourra croître. En moyenne les perturbations de ce fluide ne commenceront à croître qu'après le découplage.

- *Le découplage.* La pression du rayonnement devient rapidement inopérante : $c_s \rightarrow 0$. La longueur de Jeans tend vers 0 : tous les modes croissent, les instabilités gravitationnelles se forment et la composante baryonique tombe dans les puits de potentiels creusés par la matière sombre depuis la période d'équivalence.

Nous avons donc vu dans cette section comment, en partant de l'inflation, ont pu se former les fluctuations de densité qu'on observe aujourd'hui dans l'univers à travers les anisotropies primaires du CMB et la présence de grandes structures massives. Bien que notre approche se place dans un cadre non relativiste les résultats restent valables et permettent de décrire correctement (et simplement) la période de passage d'une ère dominée par le rayonnement vers une ère dominée par la matière non-relativiste. On peut consulter la revue de J. Bartlett (Bartlett, 1999) pour avoir le traitement relativiste de la croissance des perturbations. Dans la section suivante je présente tout d'abord ce qu'il advient du rayonnement après cette période de transition, ensuite j'introduirai le travail présenté dans le chapitre suivant concernant l'avenir de la matière après cette même transition, celle-ci jouant alors un rôle primordial dans l'évolution de l'univers.

1.3.3 Après le découplage

Après la période de découplage entre le rayonnement et la matière, présentée dans la partie précédente, on observe dans l'univers ces deux composantes qui vont évoluer de manière quasiment indépendantes. Dans un premier temps je décris le devenir du rayonnement puis celui de la matière.

1.3.3.1 Le CMB

Le CMB a été émis suite à la *recombinaison* des électrons aux noyaux pour la première fois depuis le début de l'univers, alors que la température permet à cette réaction de se produire et d'être stable. Auparavant, la température était telle que les électrons étaient sans cesse dissociés des noyaux : l'énergie de liaison de l'atome d'hydrogène étant inférieure à l'énergie thermique des photons. Cette période de recombinaison dure un certain temps, on peut montrer qu'elle s'échelonne sur la plage de *redshift* suivante :

$$1200 < z_{rec} < 1400$$

en fonction de la valeur de $\Omega_b h^2$ (entre 0,01 et 1). Pour un *redshift* intermédiaire : $z \simeq 1300$ on a la température de recombinaison $T_{rec} \simeq 0.3 \text{ eV} \simeq 3600 \text{ K}$. Le découplage survient lorsque le libre parcours moyen des photons devient plus grand que l'âge de l'univers, soit pour des *redshifts* :

$$1100 < z_{dec} < 1200$$

ce qui correspond à une température de $T_{dec} \simeq 0.25 \text{ eV} \simeq 3000 \text{ K}$. Une fois découplés les photons se propagent librement : nous les observons aujourd'hui à une température de 2.725 K . Le CMB étant une des prédictions fondamentale du modèle du Big-Bang, de nombreuses expériences ont été mises en place pour mesurer son spectre. La première expérience COBE, lancée en 1989, embarquant l'instrument FIRAS (voir introduction) a permis d'obtenir la température de corps noir du CMB. Ensuite en 1992 Smoot et al. (1992) observe les premières anisotropies grâce à l'instrument D.M.R. (toujours sur COBE). Ces anisotropies ont des tailles angulaires de l'ordre de 10° avec des dispersions de l'ordre de $30 \pm 5 \mu\text{K}$. Cette observation très importante montre que l'univers à grande échelle est homogène, ce qui valide l'hypothèse du principe cosmologique. On peut alors se poser la question suivante : comment les fluctuations de température du CMB (observées) sont elles liées aux fluctuations de densité au moment du découplage ?

Les fluctuations de température dans le CMB à toutes les échelles s'expliquent à travers le traitement relativiste de l'évolution des fluctuations de densité dans un fluide de baryons-photons. On peut montrer la présence *d'oscillations acoustiques* dans un tel fluide : on a alternativement des zones sur-denses qui deviennent sous-denses et inversement. Les fluctuations de température ont alors deux origines :

- la principal due aux fluctuations dans le potentiel gravitationnel (présentes à toutes les échelles) : les photons, pour parvenir jusqu'à nous ont été redshiftés s'ils provenaient d'une région où le potentiel gravitationnel était plus fort que la moyenne, et blueshiftés s'ils provenaient d'une région où le potentiel gravitationnel était plus faible. Ainsi ils nous apparaîtront plus chauds ou plus froids que la moyenne : ceci se traduit sur le spectre de puissance du CMB par des pics dits adiabatiques.
- Le second effet, qui se traduit sur le spectre du CMB par la présence de pic Doppler, est dû au fait qu'au moment de leur dernière diffusion les photons rencontrent des électrons animés de vitesses différentes en opposition de phase avec les variations de densité : les zones sur-denses ou sous-denses ont des vitesses nulles tandis que des zones où la densité varie auront une contribution aux pics Doppler importante.

Afin de comprendre plus clairement ce phénomène on utilise une grandeur observationnelle du CMB : son spectre de puissance. Il s'exprime habituellement à travers sa décomposition en harmoniques sphériques (la base naturelle pour les problèmes à symétrie sphérique). Ainsi les fluctuations de température du CMB sont notées :

$$\frac{\Delta T}{T} = \sum_{l=2}^{\infty} \sum_{m=-l}^l a_{lm} Y_l^m \quad (1.58)$$

où les sommes se font sur les multipôles l . A partir d'une carte de température (grandeur observée) on peut calculer les a_{lm} en utilisant l'orthogonalité des harmoniques sphériques :

$$a_{lm} = \int \frac{\Delta T}{T}(\vec{n}) Y_l^{m*}(\vec{n}) d\vec{n} \quad (1.59)$$

Le spectre de puissance est relié aux coefficients a_{lm} par :

$$C_l = \langle |a_{lm}|^2 \rangle = \frac{1}{2l+1} \sum_{m=-l}^l |a_{lm}|^2 \quad (1.60)$$

Les contraintes et estimations sur les paramètres cosmologiques se font en étudiant par exemple l'amplitude des pics dans le spectre de C_l ainsi que leur espacement absolu et relatif. Plus une expérience sera capable de voir de pics plus elle sera en mesure de fournir des contraintes fortes sur les paramètres cosmologiques. Pour cela on doit être en mesure de voir des structures de plus en plus petites, de façon à atteindre des multipôles l de plus en plus grands (l varie comme l'inverse des tailles angulaire θ). Par exemple on comprend facilement que la position des pics en fonction de l'échelle angulaire donne une information riche sur la courbure spatiale de l'univers à savoir le paramètre Ω_k ce qui revient à la somme des paramètres $\Omega_M + \Omega_\Lambda$. En effet la taille apparente des fluctuations (la position en l des pics) est directement reliée à la géométrie de l'univers : sphérique, plat ou hyperbolique.

FIRAS observait le monopôle, DMR le multipôle d'ordre 20 (dix degrés sur le ciel) tandis que BOOMERANG a relevé le premier véritable spectre du CMB jusqu'aux multipôles d'ordre 600 (de Bernardis et al., 2000; Melchiorri et al., 2000). Archeops a permis d'obtenir et de mesurer clairement l'amplitude et la position du premier pic acoustique. WMAP a fixé les choses en mesurant le spectre du CMB jusqu'à des multipôles de l'ordre de 600 en observant tout le ciel. CBI et ACBAR, deux expériences haute résolution ont mesuré le spectre du CMB aux multipôles supérieurs mais avec de grosses barres d'erreurs (Kuo et al., 2004), (Mason et al., 2003). Les prochaines expériences comme Olimpo ou APEX feront des observations jusqu'aux multipôles d'ordre 2200 avec des sensibilités jusqu'à 5 fois meilleures que BOOMERANG. Le satellite Planck quant à lui permettra une mesure plein ciel de la polarisation du CMB et une mesure du spectre de puissance des anisotropies, quasiment limitée par la variance cosmique, jusqu'aux multipôles d'ordre 2000. La figure 1.3.1 illustre l'observation des anisotropies du CMB par WMAP.

Les observations du CMB sont expérimentalement difficiles, étant donné qu'il constitue l'arrière plan théoriquement le plus lointain que l'on puisse observer (en restant dans le domaine du rayonnement : on peut supposer observer un jour un "CNB", *Cosmic Neutrino Background*, provenant de plus loin encore). Ainsi de nombreux phénomènes astrophysiques d'avant plan vont venir contaminer le signal des anisotropies primordiales :

- l'effet Sachs-Wolfe intégré dû aux variations du potentiel gravitationnel entre le CMB et l'observateur,
- les effets de weak-lensing autour des structures massives,
- la période de ré-ionisation de l'univers par le rayonnement des premières étoiles,
- l'effet SZ (sujet de cette thèse),
- les émissions ponctuelles radios, infrarouges, etc.

Quelques uns de ces effets seront présentés dans le chapitre 4, dont bien sur, l'effet SZ, pour lequel, le CMB est à la fois l'origine et un contaminant ! L'amplitude prédite des anisotropies secondaires est souvent faible par rapport aux anisotropies primaires. Cependant pour l'effet SZ l'ordre de grandeur est le même pour les multipôles $l > 2000$ que les prochaines expériences seront en mesure d'atteindre.

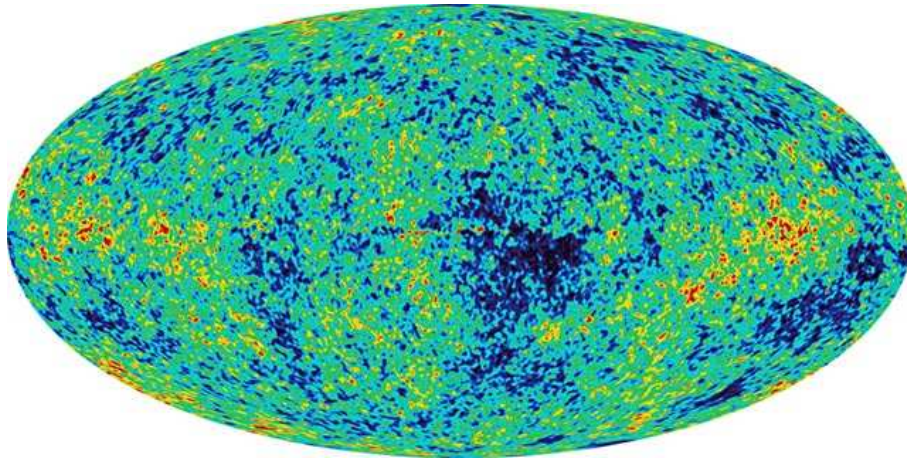


FIG. 1.3.1 – Carte plein ciel des anisotropies du CMB simulée à la résolution et la sensibilité de l'expérience WMAP.

1.3.3.2 Vers la formation des grandes structures

Dans le paragraphe précédent nous avons donc vu quel était l'avenir du rayonnement après le découplage : une dilution dans l'univers en expansion, donnant aujourd'hui le rayonnement fossile de l'univers, observé à 2.725 K. Le chapitre suivant (chapitre 2) présente l'évolution, après le découplage, des fluctuations de matière jusqu'à la formation des amas de galaxies. La figure 1.3.2 illustre l'observation des grandes structures grâce au sondage de galaxies 2dF.

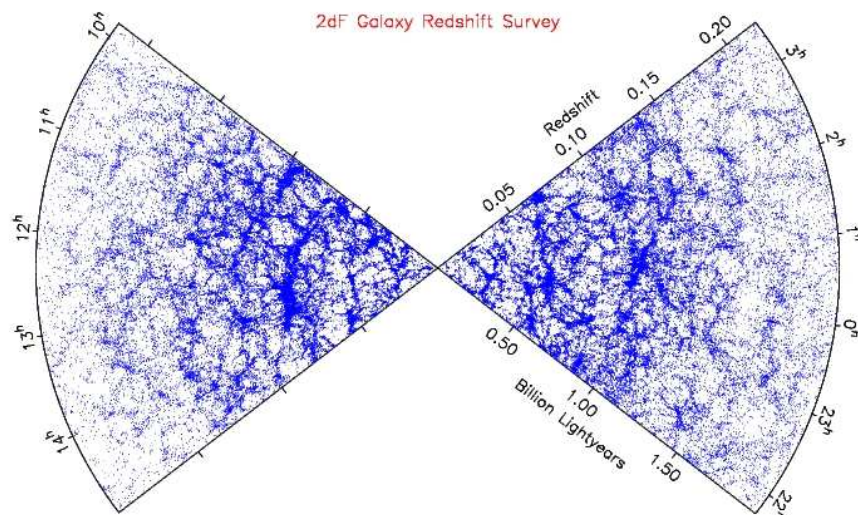


FIG. 1.3.2 – Les grandes structures vues par le sondage de galaxies 2dF.

Nous ferons dans un premier temps le point sur les observations des amas de galaxies en suivant un fil composé de deux brins : chronologique et selon la longueur d'onde de la sonde utilisée pour effectuer l'observation (optique, X, SZ et

weak-lensing). Ensuite je présenterai un modèle d'effondrement sphérique (puis son extension asymétrique) permettant de décrire les évolutions hautement non-linéaires des structures. Ceci nous permettra alors d'aborder deux descriptions des amas : tout d'abord une description "individuelle" à travers leur équilibre thermodynamique, et enfin une description statistique à travers le formalisme de la fonction de masse donnant leur abondance en fonction de leur masse et de leur *redshift*. Pour terminer j'introduirai l'effet SZ, son intérêt pour la cosmologie, le formalisme associé, et enfin comment la mesure de l'abondance des amas permet de contraindre les paramètres cosmologiques. Ces deux composantes, rayonnement et matière, séparées au moment du découplage, ayant la même origine (les fluctuations de densité) et suivant des évolutions complètement différentes, interagissent des milliards d'années plus tard à travers l'effet SZ et apportent chacune des informations complémentaires pour comprendre l'univers.

Chapitre 2

Amas de galaxies : formation, observations SZ et cosmologie

Sommaire

2.1	Les amas : observations multi-longueurs d'onde	35
2.1.1	Premières études en optique : les galaxies	35
2.1.2	L'apport des observations en X : étude du gaz	36
2.1.3	Les observations SZ pointées	36
2.1.4	Le weak lensing : sonder la matière sombre	36
2.1.5	Un amas en quelques chiffres	37
2.2	Description individuelle des amas	39
2.2.1	Le modèle sphérique d'effondrement gravitationnel	39
2.2.2	Physique de l'état lié	43
2.2.2.1	Les profils de densité	43
2.2.2.2	Calcul du rayon viriel : $R_v(M)$	45
2.2.2.3	Calcul de la température : $T(z, M)$	45
2.2.3	Effondrement asymétrique	46
2.3	Description statistique des amas : fonction de masse . .	47
2.3.1	Formalisme de Press & Schechter	47
2.3.2	D'autres fonctions de masse	50
2.3.3	Calcul de la distribution en masse et <i>redshift</i> des amas . .	50
2.4	Observations par effet SZ thermique	53
2.5	Amas de galaxies et cosmologie	57

Dans ce chapitre je décris les amas de galaxies en répondant aux questions suivantes :

- **Comment a-t-on observé les amas de galaxies et comment les observe-t-on maintenant ?**

Je répondrai en décrivant chronologiquement les observations des amas de galaxies à différentes longueurs d'onde : optique, X, SZ et weak-lensing.

- **Comment se forment-ils ?**

Je présenterai le modèle sphérique de l'effondrement gravitationnel qui permet d'introduire des grandeurs importantes intervenant dans la description individuelle et statistique des amas.

- **Quel est leur intérêt pour la cosmologie ?**

J'introduirai le formalisme de la fonction de masse qui est le lien direct entre la distribution en masse et *redshift* des amas et le cadre cosmologique.

- **Quel est l'intérêt des observations par effet SZ ?**

J'expliquerai l'effet Sunyaev Zel'dovitch thermique et soulignerai son intérêt pour étudier la cosmologie et comprendre les mécanismes de formation des amas et leur structure interne.

2.1 Les amas : observations multi-longueurs d'onde

Les premières études d'amas de galaxies dans le visible remontent aux années 1930. Leur structure sera mieux comprise dans les années 60 avec l'arrivée des observations en X qui permettent de sonder la principale composante des amas : le gaz. Aujourd'hui les amas sont des structures relativement bien comprises mais des questions restent en suspens.

La figure 2.1.1 illustre l'observation d'amas de galaxies dans différentes longueurs d'onde.

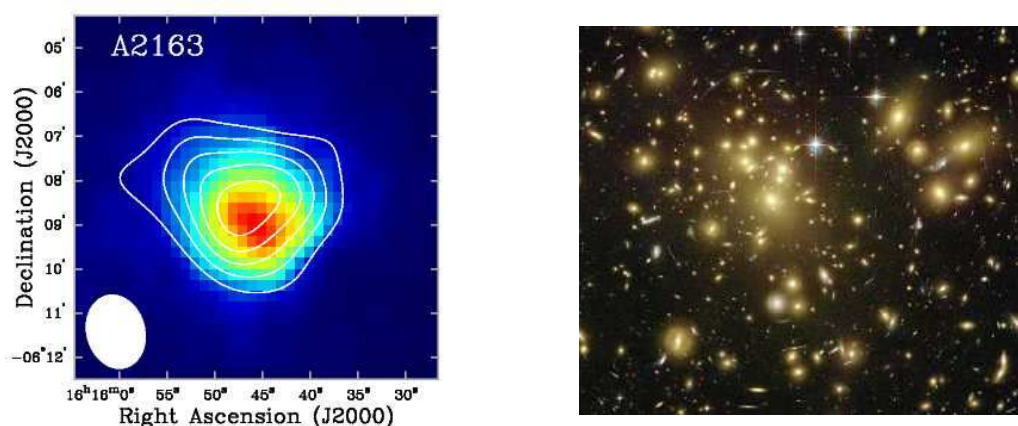


FIG. 2.1.1 – A gauche l'amas A2163 vu en rayon X (contours) et par effet SZ (couleurs). A droite un exemple d'amas de galaxies en optique. On distingue nettement une surdensité de galaxies et des effets de lentilles gravitationnelles : deux autres moyens pour observer les amas.

2.1.1 Premières études en optique : les galaxies

Les amas des années 30 sont vus comme des surdensités de galaxies : quelques centaines dans un rayon de quelques Mpc. L'amas le plus proche de nous, Coma, est observé par Zwicky. Il estime sa masse (par la somme des masses de galaxies le composant) et il estime aussi leur dispersion de vitesse. La notion de “masse manquante” est introduite : les galaxies sont trop rapides ou Coma trop léger ! L'étude systématique des amas a débuté en 1958 par l'établissement du premier catalogue (2712 sources) d'amas en optique d'Abell (Abell, 1958) puis par celui de Zwicky en 1960 comptant plus de 10000 amas. En 1962 King propose un modèle ajustant la densité d'étoiles dans un amas stellaire en fonction du rayon. Ce modèle est utilisé avec succès pour décrire le profil de densité des galaxies à l'intérieur des amas.

Aujourd'hui l'instrument CFHTLS grâce au sondage “wide” a permis de détecter une centaine d'amas (définis comme des surdensités locales de galaxies) jusqu'à des redshifts photométriques de 1,2. Une densité de 40 deg^{-2} a été mesurée.

2.1.2 L'apport des observations en X : étude du gaz

En 1966 les premières observations extragalactiques en X sont effectuées (Byram et al., 1966). Il faut attendre 1970 et le satellite Uhuru pour obtenir le premier catalogue de sources extragalactiques en X (Giacconi et al., 1972), (Giacconi et al., 1974). On constate alors que les amas sont des sources X intenses et étendues. On explique cette émission par *bremsstrahlung* : le freinage des électrons dans le champ électromagnétique des ions du gaz chaud intra-amas. Le rayonnement *bremsstrahlung* des amas est la voie principale de refroidissement du gaz intra-amas. Le temps de refroidissement typique d'un amas par ce biais est supérieur à l'âge de l'univers, l'évolution de la température du gaz, pour un amas thermalisé, est considéré comme quasi-stationnaire. Le gaz représente les quatre cinquième de la masse baryonique de l'amas. Le profil de King a été adapté pour la description de la densité du gaz dans les amas (voir le paragraphe 2.2.2.1).

Depuis ces observations, de nombreux satellites se sont relayés pour observer les amas de galaxies en X. La revue de Sarazin (1986) décrit tous les progrès réalisés depuis Uhuru jusqu'au satellite Einstein. Il faut attendre 1990 et ROSAT pour avoir un sondage complet du ciel en rayon X (plus de 120000 sources et 9 ans de prise de données) puis à sa suite, Chandra et XMM-Newton qui observent des amas pointés à haute résolution et haute sensibilité (Finoguenov et al., 2001a), (Arnaud et al., 2002), (De Grandi and Molendi, 2002), (Takahashi and Yamashita, 2003), (Arnaud et al., 2004), (Pointecouteau et al., 2004). Ces nouvelles données permettent d'accéder non seulement aux profils de densité du gaz mais aussi à ses profils de température et de métallicité. La dernière revue concernant les observations en X des amas est celle de Rosati et al. (2002).

2.1.3 Les observations SZ pointées

On peut trouver une revue des premières observations SZ pointés dans Carlstrom et al. (2001). Les résultats des expériences OVRO, SuZIE, ACBAR, CBI, BIMA, etc, y sont présentés. Ces observations ont confirmé l'intérêt de l'effet SZ pour sonder les amas lointains et comprendre leurs mécanismes de formation. En effet, l'effet SZ permet d'observer les couches externes des amas, contrairement aux observations en X qui sont plus sensibles au coeur. Les observations conjointes SZ et X permettent d'étudier les profils en température et densité des amas avec une grande précision. Par ailleurs les observations conjointes d'amas en SZ et X permette de déterminer la distance angulaire de l'amas et on peut donc espérer contraindre le paramètre h par cette méthode (Reese et al., 2002).

Une revue récente de l'effet SZ peut être trouvée dans l'article de Rephaeli (2005).

2.1.4 Le weak lensing : sonder la matière sombre

Le *weak-lensing*, ou effet de lentille gravitationnelle faible, permet de sonder directement la masse gravitationnelle des structures massives de l'univers. Les avantages d'une telle sonde apparaissent clairement : on évite la physique intermédiaire, mal comprise, entre la masse des amas et leur luminosité (optique, X ou SZ) dépendant fortement de l'histoire propre de la structure considérée. En 1991 Tyson (Tyson,

1991) observe le premier signal de weak-lensing en observant un amas proche du catalogue Abell. Dans les années qui suivent de nombreux amas proches verront leur structure massive sondée grâce aux lentilles gravitationnelles.

Les mesures de *weak-shear* ou cisaillement gravitationnel faible sur un grand champ sont extrêmement difficile à obtenir car le rapport signal sur bruit est très faible. Par ailleurs les contraintes instrumentales sont importantes notamment du fait de la PSF de l'expérience. En effet dans ces observations on cherche à mesurer les cisaillements des nombreuses galaxies observées afin d'en déduire une carte de densité permettant de détecter les zones sur-denses : les amas. Une PSF anisotrope constitue un bruit pour les mesures de weak-shear. Par ailleurs cette PSF, pour les expériences au sol, dépend du moment de l'observation en raison des déformations de l'antenne sous l'action de la gravitation selon sa direction de pointage. Tout ceci contribue à la difficulté de telles expériences qui dans l'avenir seront embarquées en satellite.

Le sondage du CFHTLS-wide mené par des équipes françaises et canadiennes a pour objectif de couvrir 150 deg^2 en trois morceaux de 50 deg^2 chacun. Ce sondage est mené dans les bandes optiques (u^* , g' , r' , i' , z'). Ses objectifs scientifiques sont l'étude des grandes structures et de la distribution de la matière dans l'univers grâce à la mesure de weak-shear et à l'observation de la distribution des galaxies. L'expérience a connu des difficultés, l'efficacité d'observation par nuit ayant été sur-estimée. Les premiers résultats concernant le weak-shear devraient être publiés cet été par Hoekstra (2005). L'expérience DUNE (pour *Dark UNiverse Explorer*) à deux objectifs principaux :

- l'observation d'amas de galaxies par la mesure de weak-shear,
- l'observation de supenovae de type Ia.

Ces deux types d'observations permettront de contraindre fortement les paramètres des modèles décrivant l'énergie noire. Afin de remplir ces objectifs et de prendre la suite du CFHTLS dans les mesures de weak-shear, DUNE sera une expérience spatiale, embarquant une caméra CCD à grand champ de vue (FOV pour *Field Of View*) 0.55 deg^2 et un télescope de 1.2 m. Cette configuration permettra une résolution angulaire à mi-hauteur (FWHM pour *Full Width of Half Maximum*) de l'ordre de $0.25''$. Le sondage est prévu pour durer trois ans avec une efficacité d'observation de 60% et couvrir 20000 deg^2 . Un suivi photométrique de mesure des redshifts sera nécessaire et devra atteindre la précision de 0.1 jusqu'aux redshifts de 1. Le projet est prévu pour démarrer en 2011.

2.1.5 Un amas en quelques chiffres

Les valeurs typiques des caractéristiques des amas sont données dans le tableau 2.1.5.

Masse totale	$10^{14} \sim 10^{16} M_{\odot}$
Matière sombre	$\sim 87\%$
Gaz	$\sim 10\%$
Galaxies et étoiles	$\sim 3\%$
Rayon	$\sim \text{Mpc}$
Température du gaz	$10^7 \sim 10^8 \text{ K}$
Densité électronique	$10^{-5} \sim 10^{-2} \text{ cm}^{-3}$

TAB. 2.1 – Caractéristiques des amas. Les pourcentages se réfèrent à la masse totale de l’amas.

Leur masse varie énormément selon le type d’amas. Les groupes sont composés de quelques galaxies, tandis que les super-amas sont composés de plus d’un millier de galaxies. Les performances croissantes des observations optiques, X et SZ, et des simulations N-corps permettent aujourd’hui d’avoir une bonne compréhension de la physique interne de ces structures. Il est possible de sonder les profils de densité des galaxies et du gaz, mais aussi le profil de la température ou de la métallicité (Finoguenov et al., 2001a), (Arnaud et al., 2004), (da Silva et al., 2004), (Pointecouteau et al., 2004), (Neumann et al., 2001). La fraction de gaz dans les amas a été étudiée par Mathiesen et al. (1999), Grego et al. (2001). Cette mesure permet d’apporter des contraintes sur le rapport des paramètres cosmologiques Ω_b et Ω_M : $f_g = \Omega_b/\Omega_M$.

2.2 Description individuelle des amas

Cette section explique comment, dans un univers en expansion, un amas peut se former à partir d'une surdensité. Ensuite, en supposant l'équilibre thermique de la structure formée, j'explique comment on peut déduire la distribution de la matière dans l'amas, la température du gaz intra-amas, sa taille, etc.

2.2.1 Le modèle sphérique d'effondrement gravitationnel

À la fin du chapitre 1 (1.3.3.2) nous avons vu qu'après le découplage de la matière et du rayonnement les surdensités croissent et deviennent de plus en plus denses. Nous utilisons alors une approche perturbative pour décrire leur évolution. Cette approche ne permet cependant pas de décrire l'effondrement gravitationnel intense, fortement non linéaire, qui suit cette période. Le modèle d'effondrement sphérique présenté ici le permet. Il est exposé en détail dans Kitayama (1997), Peacock (1999) ou encore Rich (2002).

On étudie l'effondrement sphérique d'un nuage de gaz. En nous plaçant dans un cadre Newtonien et en considérant un découpage du nuage en couronnes sphériques qui ne se chevauchent pas, on peut décrire l'évolution temporelle d'une couronne de rayon physique $r(t)$ par l'équation du mouvement suivante :

$$\frac{d^2r(t)}{dt^2} + \frac{GM}{r^2(t)} = 0 \quad (2.1)$$

On appelle M la masse totale incluse à l'intérieur de la coquille de rayon $r(t)$. On s'intéresse à l'évolution de cette structure fermée : la masse M est une constante du mouvement. Cette équation décrit l'expansion et la contraction d'une sphère de rayon $r(t)$ sous l'effet de l'autogravitation. On peut écrire que M est une constante dans le cas où des coquilles concentriques ne se chevauchent pas : il n'y a pas d'échanges de matière entre elles et l'extérieur.

On peut intégrer une première fois cette équation ($\int dr$) sous la forme :

$$\frac{1}{2} \left(\frac{dr}{dt} \right)^2 - \frac{GM}{r} = -E < 0 \quad (2.2)$$

c'est l'équation de conservation de l'énergie d'un système gravitationnellement lié avec $E > 0$. La solution à l'équation précédente qui conduit à la formation d'un système lié est la suivante :

$$\begin{cases} r = \frac{r_m}{2}(1 - \cos\theta) \\ t = \frac{t_m}{\pi}(\theta - \sin\theta) \\ \theta \in [0, 2\pi] \end{cases} \quad (2.3)$$

où on a introduit $r_m = 2\frac{GM}{E}$ et $t_m = \pi\frac{GM}{E^{3/2}}$, le rayon et le temps correspondant à $\theta = \pi$. La figure 2.2.1 illustre l'évolution du rayon d'une sphère de masse constante dans le cas d'un univers Einstein-de Sitter parfaitement uniforme et dans le cas d'une surdensité. La densité à l'intérieur de la coquille est donnée par :

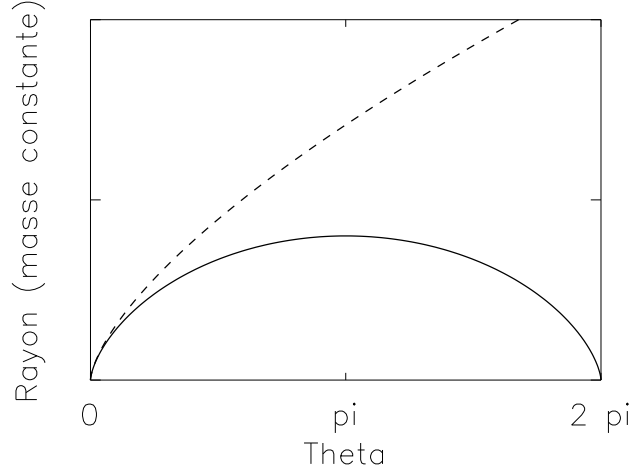


FIG. 2.2.1 – Évolution du rayon d’une sphère de masse constante dans le cas d’une surdensité (trait plein) et dans le cas d’un univers parfaitement uniforme (tirets) : $r \propto t^{2/3}$.

$$\bar{\rho}(< r, t) = \frac{3M}{4\pi r^3} \quad (2.4)$$

On cherche dans cette section à définir un critère sur la formation d’une structure virielle pour le champ des perturbations *linéaires* de densité. On regarde donc le comportement dans un régime linéaire où $\theta \ll 1$:

$$\begin{cases} r \simeq \frac{r_m}{2} \frac{\theta^2}{2} \\ t \simeq \frac{t_m}{\pi} \frac{\theta^3}{6} \end{cases} \quad (2.5)$$

On cherche alors un critère sur le contraste de densité δ défini comme :

$$\bar{\delta} = \frac{\bar{\rho}(< r, t) - \bar{\rho}(t)}{\bar{\rho}(t)} \quad (2.6)$$

On s’intéresse pour cela aux contrastes, linéaire Δ^l ou non-linéaire Δ^{nl} , définis à partir des expressions de la densité moyenne à l’intérieur de la sphère de rayon r dans le cas où on a utilisé une extrapolation linéaire $\bar{\rho}^l(< r, t)$ et dans le cas où on considère la valeur réelle de cette densité : $\bar{\rho}^{nl}(< r, t)$. En utilisant les expressions linéarisées de r et t , on obtient, par extrapolation linéaire, pour le contraste linéaire :

$$\Delta^l \equiv \frac{\bar{\rho}^l(< r, t)}{\bar{\rho}^{EdS}(t)} = 1 + \frac{3}{20} [6(\theta - \sin(\theta))]^{2/3} \quad (2.7)$$

et pour le contraste non linéaire :

$$\Delta^{nl} \equiv \frac{\bar{\rho}^{nl}(< r, t)}{\bar{\rho}^{EdS}(t)} = \frac{9 (\theta - \sin(\theta))^2}{2 (1 - \cos(\theta))^3} \quad (2.8)$$

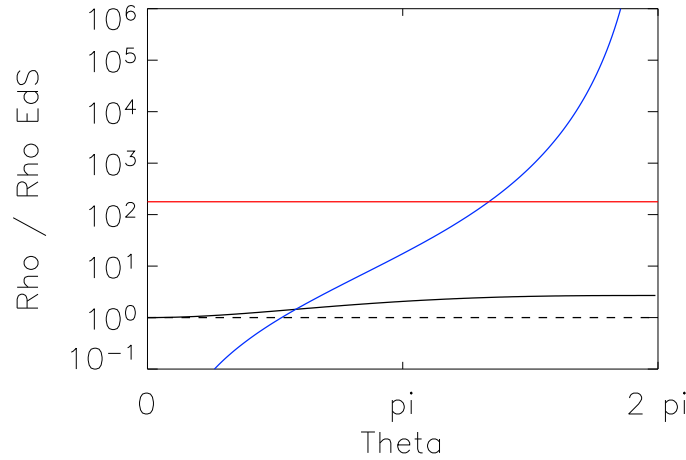


FIG. 2.2.2 – Rapport des densités, linéaire et non-linéaire, sur la densité moyenne de l’univers (ici Einstein-de Sitter). Le trait en pointillés correspond à l’évolution moyenne d’un univers uniforme. Le trait rouge correspond au système virialisé. Le trait bleu représente l’évolution non-linéaire de la surdensité, tandis que le trait continu noir représente l’extrapolation linéaire de cette évolution.

En suivant l’évolution de la sphère illustrée par les figures 2.2.1 et 2.2.2, on décompose l’effondrement en 4 étapes :

1. Expansion : dans un premier temps la sphère suit l’expansion cosmologique (courbe en pointillés de la figure 2.2.1).
2. Equilibre : les équations d’évolution montrent que pour $\theta = \pi$ le système atteint un point “de retour” à partir duquel son évolution est découplée de l’expansion de l’univers. On a alors :

$$\begin{aligned} r &= r_m = 2 \frac{GM}{E} \\ t &= t_m = \pi \frac{GM}{E^{3/2}} \end{aligned} \quad (2.9)$$

3. Effondrement : si les forces entre les coquilles de rayon r restaient radiales on aurait un effondrement complet $r = 0$ atteint en $\theta = 2\pi$. On aurait alors un contraste non-linéaire de densité infini (2.8).
4. Virialisation : Cette solution montre qu’au temps $\theta = 2\pi$ le système atteint un rayon nul avec une densité infinie. En réalité la coquille extérieure subit un phénomène de “relaxation violente”. La pression du gaz joue un rôle de force de rappel opposée à l’action de la gravitation. Les forces en présence ne sont plus uniquement radiales : la sphère atteint un rayon non nulle, le rayon *viriel* de la structure, donné par :

$$r_{vir} = \frac{r(\theta = \pi)}{2} = \frac{GM}{E} \quad (2.10)$$

On suppose que le temps auquel le système atteint cet équilibre correspond à $\theta = 2\pi$:

$$t_{vir} = 2\pi \frac{GM}{E^{3/2}} \quad (2.11)$$

On peut calculer le rapport des densités de la région sphérique et de l'univers à cette époque. Dans le modèle non linéaire on obtient :

$$\Delta_c \equiv \Delta^{nl}(r_{vir}) = \frac{\bar{\rho}^{nl}(< r_{vir}, t_{vir})}{\bar{\rho}(t_{vir})} = 18\pi^2 \simeq 178 \quad (2.12)$$

et dans l'approximation linéaire :

$$\Delta^l(r_{vir}) = \frac{3}{20}(12\pi)^{2/3} \simeq 2.69 \quad (2.13)$$

Il est important de noter que Δ est indépendant de G , M , E et de t_{vir} . Dans le cas d'univers différents de EdS on observe une dépendance en temps.

Dans les modèles décrivant les amas, individuellement ou statistiquement, on utilise ces deux quantités :

- $\delta_c \equiv \Delta^l(r_{vir}) - 1$ pour calculer l'abondance des amas d'une masse donnée et à un *redshift* donné (voir la section 2.3) et
- Δ_c pour estimer la température du gaz, le rayon viriel de l'amas, etc (voir la section 2.2.2).

Dans le formalisme de la fonction de masse (section 2.3) le critère retenu pour considérer qu'une surdensité s'est virialisée est le suivant, une structure est virialisée à un instant t si à un instant antérieur donné t_i la condition suivante a été satisfaite :

$$\delta(< r, t_i) > \delta^l(< r_{vir}, t_{vir}) = \Delta^l(r_{vir}) - 1 \equiv \delta_c \simeq 1.69 \quad (2.14)$$

Si cette condition est remplie on considère qu'au temps t il existe une structure formée à cette position. Je rappelle ici les résultats présentés par Kitayama dans sa thèse (Kitayama, 1997) concernant les valeurs de δ_c et Δ_c pour des univers différents de Einstein-de Sitter :

- univers de type Einstein-de Sitter, $\Omega_M = 1$ et $\Omega_\Lambda = 0$:

$$\begin{aligned} \Delta_c &= 18\pi^2 \simeq 178 \\ \delta_c &= \frac{3(12\pi)^{2/3}}{20} \simeq 1.69 \end{aligned} \quad (2.15)$$

- univers ouvert sans constante cosmologique, $\Omega_M < 1$ et $\Omega_\Lambda = 0$:

$$\begin{aligned} \Delta_c &= 4\pi^2 \frac{(\cosh \eta_{vir} - 1)^3}{(\sinh \eta_{vir} - \eta_{vir})^2} \\ \delta_c &= \frac{3}{2} \left[\frac{3 \sinh \eta_{vir} (\sinh \eta_{vir} - \eta_{vir})}{(\cosh \eta_{vir} - 1)^2} - 2 \right] \left[1 + \left(\frac{2\pi}{\sinh \eta_{vir} - \eta_{vir}} \right)^{2/3} \right] \end{aligned} \quad (2.16)$$

– univers de type Λ CDM, $\Omega_M < 1$ et $\Omega_\Lambda = 1 - \Omega_{tot}$:

$$\begin{aligned}\Delta_c &\simeq 18\pi^2 (1 + 0.4093 w_{vir}^{0.9052}) \\ \delta_c &\simeq \frac{3(12\pi)^{2/3}}{20} (1 + 0.0123 \log_{10} \Omega_M(t_{vir}))\end{aligned}\tag{2.17}$$

Dans ces équations on a : $\eta_{vir} = \cosh^{-1}(2/\Omega_M(t_{vir}) - 1)$ et $w_{vir} = 1/\Omega_M(t_{vir}) - 1$. L'équation 2.16 provient de Lacey and Cole (1993) et l'équation 2.17 est tirée de Nakamura et al. (1997).

2.2.2 Physique de l'état lié

Dans cette section je présente les différents modèles décrivant les amas virialisés que nous utilisons dans notre simulation : tout d'abord à travers les profils de densité baryonique et de matière noire communément utilisés, puis par la présentation de relations d'échelles donnant les expressions du rayon viriel r_v et de la température T dans les amas en fonction de leur masse M et de leur *redshift* z .

2.2.2.1 Les profils de densité

Baryons. Le premier modèle utilisé (toujours d'actualité) pour décrire le “gaz de galaxies” dans les amas est le modèle d'une sphère isotherme en équilibre hydrostatique. Il est décrit en détail par Binney and Tremaine (1987a,b). On considère un gaz parfait de particules (les galaxies), isotherme, en équilibre hydrostatique. L'équation décrivant cet état d'équilibre est la suivante :

$$\frac{d}{dr} \left(r^2 \frac{d \ln \rho(r)}{dr} \right) = - \frac{4\pi G}{\sigma^2} r^2 \rho(r)\tag{2.18}$$

où ρ est la densité au rayon r et σ la dispersion des vitesses selon l'axe radial. Une solution de cette équation est la “sphère singulière”. Cette solution diverge au centre de l'amas où le modèle prédit une densité infinie de galaxies. En 1972 King (King, 1972) a proposé une solution approchée, dite, “modèle de King”, en utilisant les variables réduites :

$$\begin{cases} \bar{\rho}(r) = \frac{\rho}{\rho_0} \\ \bar{r} = \frac{r}{r_c} \end{cases}\tag{2.19}$$

où ρ_0 est la densité au centre de l'amas, et r_c est le rayon de coeur de l'amas : $r_c = \sqrt{\frac{9\sigma^2}{4\pi G \rho_0}}$. En utilisant des conditions initiales adéquates. King trouve numériquement une solution qu'il approche par l'expression :

$$\bar{\rho}(r) = \frac{\rho(r)}{\rho_0} = \left(1 + \left(\frac{r}{r_c} \right)^2 \right)^{-\frac{3}{2}}\tag{2.20}$$

Par des observations optiques il a été possible de mesurer les profils de densité des galaxies et leur dispersion de vitesse. En utilisant cette dispersion il est possible d'estimer la température des particules "galaxies" :

$$T_{gal} = \frac{m_{gal}\sigma_{gal}^2}{k} \quad (2.21)$$

Les observations en X des amas ont par la suite permis d'estimer la température du gaz de baryons de l'amas : T_X . On définit le paramètre β de la manière suivante :

$$\beta = \frac{T_{gal}}{T_X} \quad (2.22)$$

En utilisant la température T_X typique de celle du gaz de baryons dans la solution de King à l'équation 2.18, on obtient le profil de densité de gaz baryonique suivant :

$$\rho_{gaz} \propto \rho_{gal}^\beta \propto \left(1 + \left(\frac{r}{r_c}\right)^2\right)^{-\frac{3}{2}\beta} \quad (2.23)$$

en utilisant les résultats d'observations en X sur une trentaine d'amas Jones and Forman (1984) donnent une valeur de beta comprise entre 0.6 et 0.8 : $\langle\beta\rangle = 0.65$. En ce qui concerne la valeur du rayon de coeur, la suivante $r_c = r_v/10$ est la plus utilisée, avec r_v le rayon viriel du halo dont j'explique l'origine et le calcul dans les paragraphes suivants. Ce modèle de King (2.23) est celui utilisé dans nos simulations pour décrire la distribution du gaz dans les amas générés. Cependant nous y avons ajouté la notion d'ellipticité ; ce point est détaillé au paragraphe 2.2.3.

Matière sombre. L'apparition de simulations numériques suffisamment précises permettent après les années 90 d'étudier le comportement de la matière noire. Un profil de densité (NFW) de celle-ci est donné par Navarro, Frenk et White (White et al., 1993). Ce profil est donné par :

$$\rho(r) = \rho_c \frac{\bar{\delta}}{\left(\frac{r}{r_s}\right)^p \left(1 + \frac{r}{r_s}\right)^{3-p}} \quad (2.24)$$

où ρ_c est la densité critique de l'univers, $\bar{\delta}$ un contraste de densité explicité dans la suite, et r_s est le rayon d'échelle, caractéristique de la structure. Jing and Suto (2000) ont récemment montré que p dépend de la masse considérée : pour les halos de type galactique (de une à quelques galaxies) on aurait $p \simeq 1.5$ tandis que pour les halos de type amas (plus d'une dizaine de galaxies) on a $p \simeq 1$. Dans notre simulation nous utilisons : $p = 1$. Les paramètres r_s et $\bar{\delta}$ sont fonctions de la masse M du halo. Un paramètre de compacité est utilisé pour décrire le profil au centre de l'amas, il vaut :

$$c = \frac{r_{200}}{r_s} \quad (2.25)$$

où r_{200} est le rayon de la région à l'intérieur de laquelle la densité moyenne est 200 fois celle de la densité critique de l'univers ρ_c . Il vérifie :

$$M_{200} = \Delta_c \frac{4\pi}{3} r_{200}^3 \rho_c \quad (2.26)$$

avec $\Delta_c = 200$ dans ce cas. On peut calculer les deux paramètres r_s et $\bar{\delta}$ en fonction de $M = M_{200}$ et c (Ma and Fry, 2000) :

$$r_s = \frac{1}{c} \left(\frac{3M}{800\pi\rho_c} \right)^{1/3} = \frac{1}{c} 1.63 \times 10^{-5} \frac{1}{\Omega_M^{1/3}} \left(\frac{M}{h^{-1}M_\odot} \right)^{1/3} h^{-1}\text{Mpc} \quad (2.27)$$

et

$$\bar{\delta} = \frac{200}{3} \frac{c^3}{\ln(1+c) - c/(1+c)} \quad (2.28)$$

Une estimation de la compacité est donné dans Cooray et al. (2000) par :

$$c(z, M) = a(z) \left(\frac{M}{M_\star(z)} \right)^{-b(z)} \quad (2.29)$$

avec $a(z) = 10.3(1+z)^{-0.3}$ et $b(z) = 0.24(1+z)^{-0.3}$. $M_\star(z)$ est l'échelle non-linéaire de masse à laquelle la hauteur du seuil $\nu(z, M)$ vaut 1 (voir l'équation 2.48). Cette compacité est choisie pour bien reproduire le spectre de puissance non-linéaire de densité. Elle est valide aux redshifts lointains (moins de 20% d'erreur jusqu'à $z = 1$) attendus pour les observations SZ, dans le cadre de modèles ΛCDM .

2.2.2.2 Calcul du rayon viriel : $R_v(M)$

La rayon viriel des amas est généralement choisi comme étant : $r_v = r_{200}$ où r_{200} est défini dans l'équation 2.26. C'est celui que nous choisissons dans notre simulation (décrite au chapitre 4). En suivant l'approche de Ma and Fry (2000), décrite précédemment, on a $r_v = r_{200} = cr_s$, soit :

$$r_v = \frac{1.63 \times 10^{-5}}{\Omega_M^{1/3}} \left(\frac{M}{h^{-1}M_\odot} \right)^{1/3} h^{-1}\text{Mpc} \quad (2.30)$$

Cette formule suppose $\Omega_\Lambda = 0$, $\Omega_k = 0$ et $z = 0$. Ces hypothèses sont fortes et un travail plus précis de modélisation sera nécessaire pour prendre en compte les évolutions de la physique du gaz en fonction du redshift et de la cosmologie.

2.2.2.3 Calcul de la température : $T(z, M)$

En suivant l'approche de Huterer and White (2002), on utilise le modèle suivant pour décrire la température des amas :

$$kT = kT_\star \left(\frac{M}{10^{15}M_\odot h^{-1}} \right)^{2/3} (\Delta_c E^2)^{1/3} \left(1 - 2 \frac{\Omega_\Lambda}{\Delta_c} \right) \text{ keV} \quad (2.31)$$

où Δ_c est défini dans le modèle d'effondrement sphérique comme la surdensité moyenne à l'intérieur du rayon viriel, et $E^2 = \Omega_m(1+z)^3 + \Omega_\Lambda + \Omega_k(1+z)^2$. Ce modèle tient compte du chauffage gravitationnel du gaz dans le cas d'amas isothermes. Le paramètre T_\star est un paramètre effectif qui permet de quantifier le surplus observé de chauffage "non-gravitationnel" du gaz dans les amas. Il est en bon accord avec les observations (Finoguenov et al., 2001b; Allen et al., 2001) et les simulations (Evrard et al., 2002; da Silva et al., 2004) pour les amas plus chauds que 3 keV. Le paramètre T_\star est un paramètre primordial dans l'utilisation des observations d'amas pour étudier la cosmologie (aussi bien en X qu'en SZ). En effet les incertitudes sur sa valeur sont les premières causes d'erreurs sur la détermination du paramètre cosmologique σ_8 que je présente dans la section sur la fonction de masse. Ainsi les valeurs de ces deux paramètres doivent être cohérentes entre elles. On utilise dans nos modèles :

$$T_\star \simeq 1.5 - 1.9 \quad (2.32)$$

correspondant à $\sigma_8 \simeq 0.85$ (Huterer and White, 2002).

La dépendance en masse est bien vérifiée (sur des amas proches) tandis que la dépendance en *redshift* souffre de plus grandes incertitudes.

D'autres modèles non isothermes peuvent être utilisés pour estimer la température des amas. Certains modèles considèrent des gaz bi-températures, d'autres suggèrent un équilibre adiabatique. Des études récentes (Neumann et al., 2001; Finoguenov et al., 2001a; De Grandi and Molendi, 2002; Arnaud, 2003; Pointecouteau et al., 2004; Arnaud et al., 2004) montrent clairement que la température des amas varie avec le rayon.

2.2.3 Effondrement asymétrique

Les profils décrits précédemment sont basés sur l'hypothèse d'une symétrie sphérique des amas. En suivant l'article de Cooray (2000) et de Sheth et al. (2001) nous avons tenu compte dans notre simulation de l'ellipticité des amas simulés. Ceci a deux conséquences : nous avons utilisé la fonction de masse de Sheth et Tormen (voir la section 2.3.2) pour calculer la distribution des amas en masse et redshift, et un profil de King projeté elliptique. En ce qui concerne ce dernier point, l'article de Cooray se base sur l'observation d'une cinquantaine d'amas en X dont la mesure de l'ellipticité a été faite (voir figure 2.2.3). Une ellipticité (sur le profil projeté) typique de 0.85 est observée en faisant l'hypothèse que les amas sont allongés (et non "plats"). Nous utilisons ces résultats pour calculer l'allongement des amas simulés.

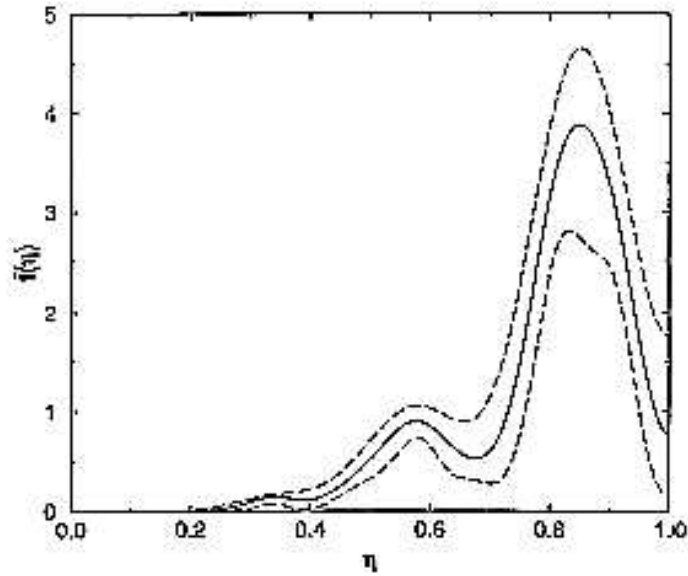


FIG. 2.2.3 – Distribution de l’ellipticité apparente d’une cinquantaine d’amas, observés en X. La ligne en tirets montre le niveau de confiance à 90% calculé par simulation Monte-Carlo.

2.3 Description statistique des amas : fonction de masse

Le modèle d’effondrement sphérique permet de comprendre la formation individuelle des amas. A partir de ce modèle et afin de décrire statistiquement la distribution en masse et redshift des amas. Press and Schechter (1974) ont introduit le formalisme de la fonction de masse, qui est décrit ici. D’abord dans le contexte général des fluctuations de densité. Puis nous présentons trois fonctions de masse : celle, initiale, de Press et Schechter, la seconde de Sheth, Mo et Tormen (Sheth et al., 2001), prenant en compte des effondrements asymétriques possibles pour les amas et s’appuyant sur des simulations numériques et la dernière de Jenkins et al. (2001) entièrement basée sur les simulations numériques précises du Hubble volume du consortium VIRGO.

2.3.1 Formalisme de Press & Schechter

Je décris ici le contexte général du modèle statistique donnant la distribution des amas en fonction de leur masse M et de leur *redshift* z . On se place juste après la période du découplage rayonnement-matière. Les surdensités croissent et se découplent de l’expansion cosmologique. Le formalisme décrit ici permet, à partir d’une description linéaire des fluctuations du champ de densité, de prédire l’abondance des structures massives (fortement non-linéaires!) de l’univers.

On décrit le contenu de l’univers par un champ de densité $\rho(\vec{x}, t)$. Une approche

en perturbation de ce champ est justifiée à une époque proche du découplage, et confirmée par les observations des anisotropies du CMB. On introduit la quantité :

$$\delta(\vec{x}, t) \equiv \frac{\rho(\vec{x}, t) - \bar{\rho}(t)}{\bar{\rho}(t)} \quad (2.33)$$

Dans un contexte d'évolution linéaire il est pratique d'utiliser la décomposition de Fourier spatiale de cette quantité : $\delta_k(t)$. Le spectre de puissance des perturbations de densité est :

$$P(k, t) \equiv \langle |\delta_k(t)|^2 \rangle \quad (2.34)$$

Ce spectre est modelisé par l'expression suivante, donnée dans Bardeen et al. (1986),

$$P(k) = k^n T(k)^2 \quad (2.35)$$

où $T(k)$ est une expression polynomiale de $q = \frac{k}{h\Gamma}$, avec n donnant la pente du spectre de puissance primordial, souvent prise égale à 1 (Spergel et al., 2003) et Γ est la pente du spectre aujourd'hui dont la valeur vaut $h\Omega_M$. On utilise une version des perturbations de densité "lissée" sur une échelle R donnée afin de définir des régions de masse $M = \frac{4}{3}\pi R^3 \bar{\rho}$:

$$\delta(M, \vec{x}, t) = \int \delta(y, t) W_M(|\vec{x} - \vec{y}|) d\vec{y} \quad (2.36)$$

$$\delta(M, \vec{x}, t) = \frac{1}{(2\pi)^3} \int \delta_k(t) \hat{W}_M(k) \exp(-i\vec{k} \cdot \vec{x}) d^3k \quad (2.37)$$

où le fonction de lissage spatial $W_M(r)$ est généralement une fonction "top hat" :

$$W_M(r) = \begin{cases} 3/(4\pi R^3) & r < R \\ 0 & r > R \end{cases} \quad (2.38)$$

et $\hat{W}_M(k)$ est sa transformée de Fourier. Ainsi définie, la variance des fluctuations de densité sur échelle de masse M sera :

$$\sigma^2(M, t) \equiv \langle |\delta(M, \vec{x}, t)|^2 \rangle = \frac{1}{(2\pi)^3} \int P(k, t) \hat{W}_M^2(k) d^3k \quad (2.39)$$

On peut obtenir une normalisation du spectre de puissance grâce à une mesure aujourd'hui de cette grandeur σ . En pratique on mesure $\sigma_8 \equiv \sigma(R = 8h^{-1} \text{ Mpc}, t_0)$. Ce paramètre, fortement dégénéré avec le paramètre T_* décrivant le chauffage du gaz intra-amas, est fondamental dans le calcul et la mesure de la distribution en masse et redshift des amas (voir le chapitre 7). On obtient ensuite facilement le spectre de puissance à un instant t donné par la relation :

$$P(k, t) = P(k, t_0) \frac{D^2(t)}{D^2(t_0)} \quad (2.40)$$

et

$$\sigma(M, t) = \sigma(M, t_0) \frac{D(t)}{D(t_0)} \quad (2.41)$$

avec $D(t)$ le facteur de croissance linéaire. En supposant que ce champ de densité possède une distribution gaussienne (prédiction de l'inflation) on peut écrire, comme l'ont supposé Press et Schechter, la probabilité d'avoir en un point, au temps t_i une surdensité $\delta(M, t_i)$:

$$P[\delta(M, t_i)] = \frac{1}{\sqrt{2\pi\sigma^2(M, t_i)}} \exp\left[-\frac{\delta^2(M, t_i)}{2\sigma^2(M, t_i)}\right] \quad (2.42)$$

La condition pour qu'une région sphérique se virialise à un temps t est que cette région ait eu à une époque précédant t_i un contraste de densité supérieur au contraste linéaire critique δ_c :

$$\delta(t_i) > \delta_c(t, t_i) \equiv \delta_c(t) \frac{D(t_i)}{D(t)} \quad (2.43)$$

Le temps t_i est défini à partir du moment où le spectre de puissance est fixé et que les perturbations commencent à croître linéairement indépendamment les uns des autres.

Cette condition permet d'écrire la fraction de la masse totale de l'univers contenue dans des structures de masse supérieure à M sous la forme :

$$F(> M, t) = \int_{\delta_c(t, t_i)}^{\infty} P[\delta(M, t_i)] d\delta \quad (2.44)$$

quel que soit t , $\delta(M, t_i)$ étant le champ de densité initial lissée sur une échelle de masse M . En effet on considère qu'une région avec une surdensité qui satisfait la condition de virialisation au temps t_i formera au temps t une structure virialisée de masse au moins égale à M . En considérant le champ de densité de distribution gaussienne précédente on peut calculer explicitement la fonction $F(> M, t)$ et sa dérivée partielle par rapport à M :

$$F(> M, t) = \int_{\delta_c(t)}^{\infty} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\delta^2/2\sigma^2} d\delta \quad (2.45)$$

où on a réécrit $\delta_c(t, t_i)$ en utilisant le fait que l'on parle de densité critique *linéaire*. En choisissant $t_i = t_0$ et en normalisant le croissance linéaire $D(t_0) = 1$ on peut écrire :

$$\delta_c(t) \equiv \delta_c(t, t_0) = \delta_c(t_0) \frac{D(t_0)}{D(t)} = \frac{\delta_c}{D(t)} \quad (2.46)$$

Avec ce choix on peut réécrire $\sigma(M, t_i)$ simplement : $\sigma(M) = \sigma(M, t_0)$. Le calcul de $\partial F / \partial M$ se fait simplement à partir de l'équation 2.45 :

$$\frac{\partial F}{\partial M} = \frac{\partial F}{\partial \delta_c} \frac{d\delta_c}{d\sigma} \frac{d\sigma}{dM} \quad (2.47)$$

En posant :

$$\frac{d\delta_c(z)}{d\sigma(M, 0)} = \nu_c(M, z) \quad (2.48)$$

on obtient :

$$\frac{\partial F}{\partial M} = \frac{1}{\sqrt{2\pi}} \frac{1}{\sigma} \nu_c e^{-\nu_c^2} \frac{d\sigma}{dM} \quad (2.49)$$

La fonction de masse utilisée par Press et Schechter est alors la suivante :

$$f_{PS} = \frac{\partial F}{\partial M} = \frac{1}{\sqrt{2\pi}} \frac{1}{\sigma} \nu_c(M, z) \left| \frac{d\sigma(M, 0)}{dM} \right| e^{-\nu_c^2/2} \quad (2.50)$$

avec $\nu_c(M, z) = \delta_c/\sigma(z, M)$ où δ_c est la contraste de densité critique nécessaire à l'effondrement de la masse M et $\sigma(z, M)$ l'amplitude moyenne des perturbations à l'échelle M et au redshift z .

2.3.2 D'autres fonctions de masse

En utilisant un modèle d'effondrement elliptique, Sheth, Mo et Tormen (Sheth et al., 2001) montrent que la fonction suivante s'ajuste aux simulations numériques de manière satisfaisante :

$$f_{ST}(z, M) = A \sqrt{2a/\pi} [1 + (1/a\nu_c^2)^p] \nu_c \left| \frac{d\ln\sigma}{d\ln M} \right| e^{-\frac{1}{2}a\nu_c^2} \quad (2.51)$$

avec $A = 0,3222$, $a = 0,707$ et $p = 0,3$. Cette fonction ne souffre pas du "problème" de la fonction de masse de Press et Schechter, à savoir une surabondance de structures de petites masses ($< 10^{14} M_\odot$) et trop peu de structures de grandes masses ($> 10^{15} M_\odot$).

Par la suite l'article Jenkins et al. (2001), s'appuyant sur des simulations du consortium VIRGO, présente la fonction suivante donnant de très bons résultats en accords avec les observations :

$$f_{VIRGO}(z, M) = 0.315 e^{-|\ln(\sigma^{-1})+0.61|^{3.8}} \left| \frac{d\ln\sigma}{d\ln M} \right| \quad (2.52)$$

2.3.3 Calcul de la distribution en masse et *redshift* des amas

La fraction de la masse totale de l'univers contenue dans des structures dont la masse est entre M et $M + dM$ est simplement :

$$\left| \frac{\partial F(> M, t)}{\partial M} \right| dM \quad (2.53)$$

La masse totale des structures de masse comprise entre M et $M + dM$ dans un volume comobile dV_0 est :

$$\left| \frac{\partial F(> M, t)}{\partial M} \right| dM \rho_0 dV_0 \quad (2.54)$$

où $\rho_0 dV_0$ est la masse comobile, c'est à dire la masse moyenne comprise dans un volume comobile : c'est une constante au cours du temps qui vaut, aujourd'hui $\Omega_M h^2 = 11.88 \times 10^{-29} \text{ g/cm}^3$. Comme $n(M, t)M$ est la masse totale des structures de masse comprise entre M et $M + dM$ dans un volume comobile dV_0 , on a :

$$n(M, t)M = \left| \frac{\partial F(> M, t)}{\partial M} \right| dM \rho_0 dV_0 \quad (2.55)$$

La fonction de masse s'écrit alors :

$$\frac{dN(M, t)}{dM dV_0} = \frac{\rho_0}{M} \left| \frac{\partial F(> M, t)}{\partial M} \right| \quad (2.56)$$

Afin de satisfaire la condition de normalisation :

$$\int_0^\infty M \frac{dN(M, t)}{dM dV_0} dM = \rho_0$$

Press et Schechter ont introduit "à la main" un facteur 2 dans la fonction de masse finale :

$$\frac{dN(M, t)}{dM dV_0} \equiv 2 \frac{\rho_0}{M} \left| \frac{\partial F(> M, t)}{\partial M} \right|$$

Ce facteur 2 est expliqué dans une approche plus rigoureuse du problème faite par Peacock & Heavens (1990) et Bond et al. (1991) : c'est le "cloud-in-cloud problem".

Le calcul du volume dV_0 est fait de la manière suivante : on observe un volume physique $dV_p = dx dy de$, où e est l'épaisseur et $dx dy = dS$ est la surface que l'on observe. Ce volume est éloigné de nous d'un redshift z . Il s'écrit en coordonnées comobiles :

$$dV_0 = \frac{1}{a(z)^3} dV_p = (1+z)^3 dS de \quad (2.57)$$

On peut relier la surface de la face observable à l'angle solide sous laquelle on la voit grâce à la distance angulaire :

$$dS = d_A^2 d\Omega \quad (2.58)$$

Par ailleurs, on écrit l'épaisseur de du volume élémentaire sous la forme $de = c dt$. On peut alors relier cette quantité à une profondeur en unité de redshift en connaissant la dépendance $t(z)$:

$$dV_0 = (1+z)^3 d_A^2 c \left| \frac{dt}{dz} \right| dz d\Omega \quad (2.59)$$

celle-ci est donnée par :

$$t(z) = \int_0^z \frac{du}{H(u)(1+u)} \quad (2.60)$$

ce qui donne : $|dt/dz| = 1/(H(z)(1+z))$ et

$$dV_0(z) = \frac{(1+z)^2 d_A^2 c}{H(z)} dz d\Omega \quad (2.61)$$

La densité d'amas calculée avec la fonction de masse de Press & Schechter est alors la suivante :

$$\frac{dN(M, z)}{dM dz d\Omega} = \sqrt{\frac{2}{\pi}} \frac{cd_A^2(1+z)^2 \bar{\rho}_0}{M^2 H(z) \sigma(M, 0)} \nu_c(M, z) \left| \frac{d\sigma(M, 0)}{dM} \right| e^{-\nu_c^2/2} \quad (2.62)$$

Le formalisme de la fonction de masse, à l'origine reposant sur des *a priori* assez fort, décrit bien les abondances d'amas observées pour des masses supérieures à $10^{13} M_\odot$ et des redshifts jusqu'à 2. La fonction de masse que nous utiliserons dans notre simulation est celle de Sheth & Tormen dont les paramètres sont évalués à partir de simulations numériques et le formalisme est bien compris en considérant des effondrements de structures asymétriques. Ce formalisme permet une mise en place légère et rapide (comparée aux simulations à N corps) de simulations de cartes d'effet SZ.

2.4 Observations par effet SZ thermique

Comme nous l'avons vu dans l'introduction de ce chapitre, les amas de galaxies sont composés d'une large fraction de gaz dont l'extension spatiale va au delà de celle des galaxies composant l'amas. Ce gaz est chaud : $T \simeq 5 \text{ keV}$ et en partie ionisé. Les photons du CMB, aujourd'hui froids, diffusent par effet Compton sur les électrons du gaz intra-amas : c'est ce que l'on appelle l'effet Sunyaev Zel'dovich (Zeldovich and Sunyaev, 1969; Sunyaev and Zeldovich, 1970, 1972, 1980a,b). Pour ces énergies la section efficace de la diffusion Compton est celle de Thomson σ_T : la profondeur optique du processus ($\int_{los} n_e \sigma_T dl$) de diffusion est très faible, de l'ordre de 10^{-4} . L'effet SZ résulte de l'interaction des deux composantes de l'univers, rayonnement et matière, séparées depuis le découplage.

Il existe plusieurs effets SZ, tous dus à une diffusion Compton des photons du CMB sur les électrons du gaz intra-amas, mais ils n'ont pas la même origine physique. On distingue donc :

- l'effet SZ thermique, le plus important, présenté et étudié dans cette thèse,
- l'effet SZ cinétique, le second en amplitude (dix fois plus faible que l'effet SZ thermique), dû au mouvement particulier de l'amas en direction de l'observateur, dans le référentiel du CMB,

et des effets SZ de plus faible amplitude (inaccessible dans les prochains sondages d'amas grand champ) comme :

- l'effet SZ polarisé, permettra de mesurer la vitesse transverse du mouvement de l'amas dans le référentiel du CMB,
- ou encore l'effet SZ des “cavités” (Colafrancesco et al., 2003; Colafrancesco, 2005), dû à la pression particulière du gaz d'électrons dans les zones sous-denses des amas,
- ou d'annihilation de particules de matière noire froide (les WIMPS en particulier) Colafrancesco (2004).

Birkinshaw fait la revue de l'effet SZ en général (Birkinshaw, 1999).

L'effet SZ thermique est dû à l'agitation thermique des électrons du gaz intra-amas. La température électronique des amas peut atteindre plusieurs keV. C'est l'effet SZ principal : son amplitude est comparable à celle des anisotropies primaires du CMB, mais contrairement à celles-ci c'est un effet localisé dans la direction des amas. Pour la plupart des amas l'effet SZ est produit par une population d'électrons non relativistes : nous nous plaçons dans ce cadre dans la suite de notre étude. L'approximation non relativiste de l'effet SZ est adaptée à des amas de température $T_e < 10 \text{ keV}$, c'est à dire pour des amas de masse moyenne ($< 10^{15} M_\odot$) en équilibre thermique. Les populations d'électrons relativistes ne sont pas négligeables dans le cas d'amas très massifs ou n'ayant pas encore atteint un équilibre thermique. L'approximation non relativiste permet d'utiliser une forme analytique pour le calcul de l'amplitude de l'effet SZ. Elle se justifie par le fait que, dans les sondages SZ grand champ comme Olimpo, on cherche à éviter les amas massifs car ceux-ci masqueraient, par leur brillance, des amas lointains, plus intéressants pour la cosmologie. Ainsi, dans notre simulation, nous nous intéresserons à une population d'amas de masse moyenne, thermalisés.

L'effet SZ thermique possède une signature spectrale caractéristique que les expériences d'observation SZ bolométriques cherchent à exploiter au maximum.

Cette dépendance, dans le cadre d'une population d'électrons non relativistes est calculée grâce à l'équation de Kompaneets qui décrit la variation du nombre d'occupation N des photons au cours de la diffusion Compton. Tous les calculs sont donnés dans la revue de Birkinshaw (1999). On obtient pour un amas :

$$\frac{\Delta I_\nu}{I_0} = y \frac{x^4 e^x}{(e^x - 1)^2} \left[\frac{x(e^x + 1)}{e^x - 1} - 4 \right] \quad (2.63)$$

avec

$$x = \frac{h\nu}{kT_{CMB}} \quad (2.64)$$

la fréquence réduite et

$$I_0 = \frac{2(kT_{CMB})^3}{(hc)^2} \quad (2.65)$$

dont l'unité est le $\text{Wm}^{-2}\text{Hz}^{-1}\text{sr}^{-1}$. La loi d'émission du corps noir est donnée par :

$$I_\nu = I_0 \frac{x^3}{e^x - 1} \quad (2.66)$$

Dans l'équation (2.63) y est le paramètre de Compton, donné par :

$$y = \int_{los} \frac{kT_e}{m_e c^2} n_e(r) \sigma_T dl \quad (2.67)$$

où l'intégrale est calculée le long de la ligne de visée (los pour *line of sight*). T_e est la température électronique de l'amas, et m_e la masse de l'électron. n_e est la densité électronique de l'amas au point r , tandis que σ_T est la section efficace de Thomson. Le paramètre y une grandeur sans unité, qui quantifie la quantité d'énergie transférée lors de la diffusion Compton.

Ainsi on peut factoriser l'effet SZ thermique non relativiste en deux parties indépendantes : la première décrivant le comportement spectral de l'effet SZ (fonction de x) et la seconde, y , dépendant uniquement des caractéristiques physiques des amas.

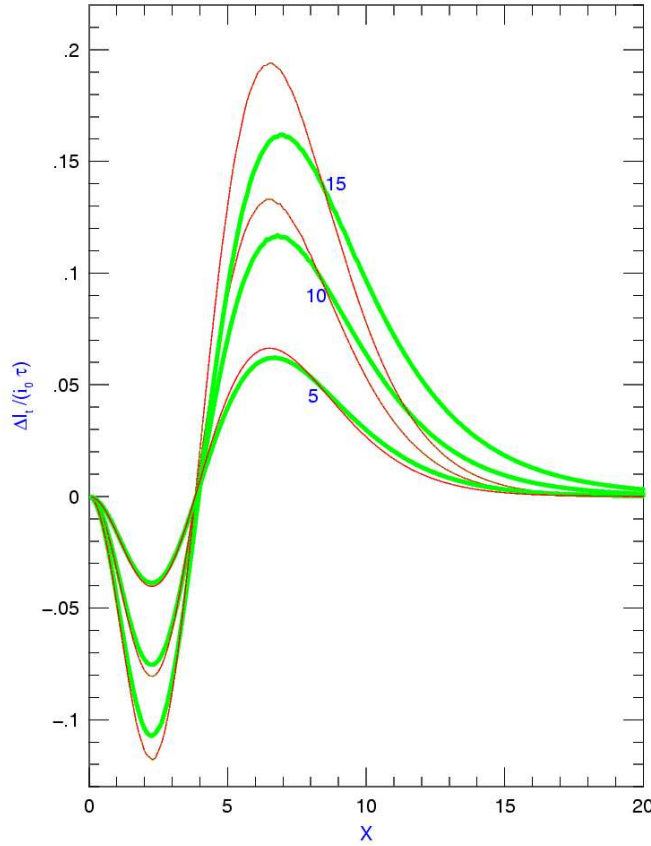


FIG. 2.4.1 – Cette figure, tirée de Rephaeli (2002b), montre pour différentes température du gaz intra-amas (5, 10 et 15 keV) le spectre de l’effet SZ thermique relativiste (courbes épaisses vertes) et non-relativiste (courbes fines et rouges).

On remarque que l’effet SZ est négatif pour $\nu < 217$ GHz, nul pour $\nu = 217$ GHz et positif pour des fréquences d’observation plus élevées. Cette signature spectrale caractéristique est exploitée par les expériences d’observation SZ multi-fréquences afin de pouvoir identifier les amas et les séparer des autres sources émettant dans cette bande de fréquence ; par exemple l’expérience Olimpo (Planck, SPT, etc.) utilisera quatre fréquences : 143, 217, 385 et 600 GHz pour détecter l’effet SZ. Si on considère une population d’électrons thermiques mais relativistes on peut montrer que la dépendance spectrale varie avec la température électronique de l’amas : le point “zéro” où l’effet SZ s’annule varie avec T_e . Comme nous l’avons précisé plus haut, nous resterons ici dans l’approximation des populations d’électrons non relativistes.

Il est important pour la suite de remarquer que l’effet SZ sur la ligne de visée (paramètre de Compton sur la ligne de visée : y) est indépendant du *redshift* de l’amas diffusant. Cette propriété rend l’effet SZ très intéressant pour les observations d’amas lointains. Un amas résolu sera bien détecté à grand *redshift*. L’effet géométrique de

la diminution de la taille angulaire de l'amas avec le *redshift* associé au fait que les expériences n'ont pas une résolution angulaire infinie fait que les amas lointains ne seront néanmoins pas observables. Cependant on s'attend à observer plus facilement des amas lointains en SZ qu'en X.

Il est intéressant de calculer le paramètre de Compton d'un amas intégré sur toute sa surface S :

$$Y = \int_S y d\Omega = \frac{k\sigma_T}{m_e c^2} \int_S \int_{los} T_e n_e(r) dl d\Omega \quad (2.68)$$

le terme $T_e n_e$ se révèle être la pression de l'amas au point r . Dans l'hypothèse d'amas isothermes (hypothèse utilisée dans notre simulation) on a $\int_S \int_{los} n_e dl d\Omega \propto M_{gaz}$ et on constate alors que le paramètre de Compton Y est directement proportionnel à la masse baryonique de l'amas M_{gaz} :

$$Y = \frac{k\sigma_T}{m_e c^2} \frac{1}{D_a^2} \frac{1 + 2/\mu}{5m_p} T_e M_{gaz} \quad (2.69)$$

où μ est le poids moléculaire moyen qui dépend de la composition du gaz, par exemple pour un amas composé de 75% d'hydrogène et 25% d'hélium on a $\mu \sim 0.6$. m_p est la masse du proton, D_a la distance angulaire et M_{gaz} la masse du gaz dans l'amas. En introduisant le paramètre f_g , la fraction de gaz, on peut écrire :

$$M_{gaz} = f_g \frac{\Omega_b}{\Omega_M} M \quad (2.70)$$

et le paramètre de Compton intégré est alors proportionnel à la masse totale de l'amas :

$$Y = \frac{k\sigma_T}{m_e c^2} \frac{1}{D_a^2} \frac{1 + 2/\mu}{5m_p} f_g \frac{\Omega_b}{\Omega_M} T_e M \quad (2.71)$$

L'effet SZ se trouve être une excellente sonde pour mesurer l'énergie totale des amas. On peut imaginer qu'il devienne possible de déterminer avec précision les limites d'un amas et donc d'estimer correctement sa masse connaissant son redshift : le paramètre de Compton intégré est un traceur de la masse totale de l'amas, indépendamment du profil de densité de celui-ci. C'est un point fort du SZ par rapport aux observations en X pour lesquelles la dépendance en brillance de surface est donnée par $\int_{los} T^{1/2} n_e^2$. La relation entre flux total X et masse est beaucoup plus complexe qu'avec l'effet SZ.

2.5 Amas de galaxies et cosmologie

L'étude des amas est un outil clef pour la compréhension de l'histoire de l'univers au même titre que l'étude des anisotropies primaires du CMB ou que les observations de supernovae de type *Ia*. Ces trois domaines sont complémentaires et apportent chacun un éclairage différent sur la forme de l'univers, son contenu et son évolution.

Méthode directe On peut, à partir de la mesure du rapport masse-luminosité $\frac{M}{L}$ dans les amas et en les comparant à une mesure moyenne pour l'univers, estimer la valeur de Ω_M . Carlberg et al. (1996) donnent la valeur suivante : $\Omega_M = 0,24 \pm 0,05 \pm 0,09$ avec ses erreurs statistiques (1σ) et systématiques. Une autre méthode consiste à calculer la fraction de baryons dans l'amas, en utilisant des observations X pour estimer la masse des baryons et une méthode pour estimer la masse totale (dispersion de vitesse, lentille gravitationnelle, ou même observation en X, via un modèle d'équilibre hydrostatique), pour en tirer la valeur de Ω_M en supposant la valeur de Ω_b connue (nucléosynthèse primordiale) (Mathiesen et al., 1999), (Roussel et al., 2000), (Grego et al., 2001). Une revue des méthodes utilisées pour contraindre Ω_M à partir des amas est donnée dans Bahcall et al. (1999).

Fonction de corrélation à deux points L'observation en X des amas par les expériences en cours permet d'estimer la fonction de corrélation à deux points actuelle. Cette fonction de corrélation apporte des informations sur l'amplitude et la forme du spectre de puissance des fluctuations de densité aux échelles des amas ($\simeq 10$ Mpc). Ce spectre est une donnée importante pour la cosmologie car c'est une quantité fondamentale pour caractériser les scénarios de formation des structures. Par ailleurs cette mesure apporte des informations, non seulement sur la forme des fluctuations de densité mais aussi sur leur origine, ce qui est en rapport avec la nature de la matière sombre. L'évolution de la fonction de corrélation en fonction du *redshift* est aussi à l'étude (Arnouts et al., 2002), (Diaferio et al., 2003), (Mei and Bartlett, 2003). Les amas sont des structures privilégiées pour ces études : le lien entre leur luminosité, X ou SZ, et leur masse est directe alors que pour les galaxies ce lien passe par la physique stellaire. Cependant des sondages comme le 2dF et SDSS avec plus de 250.000 galaxies observées contiennent une information très riche sur la cosmologie via la mesure du spectre de puissance de densité.

Dans cette thèse, nous n'étudions pas la fonction de corrélation. Nous nous limitons à l'étude des comptages d'amas. Nous étudions ces comptages en fonction de la luminosité SZ et du *redshift* des amas. Je détaille ce point dans le paragraphe suivant.

Comptages d'amas La mesure de la distribution des amas en fonction du *redshift* est une donnée importante pour la cosmologie et notamment pour la mesure de Ω_M (Barbosa et al., 1996), (Blanchard and Bartlett, 1998). L'abondance des amas de galaxies en fonction de leur masse M et de leur *redshift* z est directement reliée à la fonction de masse elle même dépendante des paramètres cosmologiques. L'abondance des amas de galaxies est ainsi fonction de plusieurs paramètres.

- Tout d’abord les paramètres cosmologiques donnant le cadre général de l’expansion de l’univers : Ω_M , la densité de matière, Ω_Λ la constante cosmologique, Ω_b la densité de baryons, h la constante de Hubble, etc.
- Ensuite les paramètres liés à la formation des structures : σ_8 , la normalisation du spectre de puissance, n la pente du spectre de puissance primordial, et la modélisation de la fonction de masse (Press & Schechter, Sheth & Tormen, Jenkins & al., etc.).
- Enfin les paramètres de la physique tardive des amas liés aux modèles décrivant l’équilibre thermique des amas : principalement T_\star , la normalisation de la relation masse-température et f_g la fraction de gaz dans l’amas.

L’effet SZ est une sonde puissante donnant une approche différente et complémentaire de l’observation en X ou optique des amas de galaxies. En effet les expériences de plus en plus performantes permettront d’apporter beaucoup d’information sur la cosmologie quand on sait que l’effet SZ permet de détecter des amas dans une large gamme de redshift et donc de sonder le passé de l’univers. Nous verrons que dans le cadre des observations SZ, où l’on n’a pas accès directement à la masse des amas, les paramètres de physique tardive ont une importance capitale (Battye and Weller, 2003) pour relier correctement la mesure de la distribution des amas à la cosmologie. En effet la luminosité SZ est directement liée à la température de l’amas elle-même liée à sa masse par une relation faisant intervenir le paramètre T_\star (2.4). L’étude des possibilités des observations SZ grand champ pour contraindre la cosmologie *via* l’abondance des amas est faite en détail dans le chapitre 7 de cette thèse.

Etant donnée la qualité des mesures des anisotropies primaires du CMB (polarisés et non polarisés) pour contraindre les paramètres cosmologiques, associées aux observations de supernovae Ia, on peut prendre comme a priori connu les paramètres cosmologiques. Je présente dans le chapitre suivant une revue rapide des différents sondages SZ et plus généralement des sondages de grandes structures, en cours de prise de données ou en cours d’élaboration. Ces grands sondages d’amas en SZ permettront de contraindre statistiquement les paramètres concernant la physique tardive des amas comme le chauffage du gaz (où les galaxies et même les étoiles jouent un rôle) : à travers son amplitude, sa variation avec le *redshift* ou encore le type d’équilibre thermique présent dans l’amas. Ces sondages associés à des observations pointées à haute résolution multi-longueurs d’onde (SZ, X, optique, Weak-Lensing) permettront de comprendre la formation des structures depuis les amas jusqu’aux étoiles qui les composent, et les phénomènes responsables du chauffage du gaz. Dans un amas, des objets astrophysiques de tailles diverses sont liés et interagissent : les amas sont à la limite entre la cosmologie et l’astrophysique et “subissent” les influences de ces deux contextes.

Chapitre 3

L'expérience Olimpo : contexte et instrument

Sommaire

3.1	Contexte	60
3.2	L'instrument Olimpo	63
3.2.1	Telescope	63
3.2.2	Cryostat	63
3.2.3	Bolomètres	63
3.2.4	Système de pointage	66
3.3	Comparaison avec APEX et Planck	67

3.1 Contexte

Ce chapitre introduit le contexte de l'expérience Olimpo en présentant, d'une manière générale les expériences SZ en cours et à venir et d'autre part en proposant une revue rapide des sondages d'amas multi-longueurs d'onde. Dans un second temps je présente les caractéristiques de l'expérience Olimpo utiles à la construction du modèle statistique de l'instrument présenté dans les chapitres 4 et 5.

L'expérience Olimpo fait partie de la troisième génération d'expériences ayant pour objectif l'observation du CMB. Après les premières observations du CMB, de ses anisotropies puis la mesure des premiers pics du spectre de puissance, les expériences actuelles cherchent à observer, soit le spectre de puissance à haute résolution angulaire, et/ou à mesurer la polarisation du CMB. Grâce à la sensibilité des détecteurs mis en place, ces expériences de troisième génération ont aussi comme objectif scientifique l'étude des avant-plan du CMB, dont les amas de galaxies observés par effet Sunyaev-Zel'dovitch.

Une revue des premières observations SZ sur des amas pointés est proposée dans Carlstrom et al. (2001) notamment les expériences du Ryle Telescope, OVRO, BIMA, SuZIE, ACBAR, CBI. Le tableau 3.1.1 illustre les principaux sondages SZ en cours ou à venir. La première série concerne les expériences à haute résolution pour les observations d'amas pointés, la seconde série présente les expériences visant aussi des observations grand champ. On peut trouver plus de renseignements sur ces expériences sur leurs sites web respectifs.

La prochaine génération d'instruments d'observation des amas par effet SZ permettra d'effectuer des comptages d'amas en fonction du *redshift* sur de grandes surfaces du ciel. Les sondages SZ comme ceux d'Olimpo, d'APEX, de Planck, combinant une bonne résolution angulaire, un champ large ($> 100 \text{ deg}^2$) et une bonne sensibilité permettront de détecter plus de 1000 amas à des redshifts intéressants pour effectuer des comptages $N(z)$. Aujourd'hui, en X, les expériences comme Chandra ou XMM ne sont pas adaptées : leur objectif est l'observation pointée d'amas proches. Cependant, des travaux sont en cours pour étudier avec XMM-LSS un catalogue d'au moins cinquante amas avec un suivi *redshift* ($z < 1.2$). Le tableau 3.1.2 donne un récapitulatif (non exhaustif) des différentes expériences de sondage multi-longueurs d'ondes d'amas.

Nom	Type	Fréquences [Ghz]	Couverture [deg ²]	FWHM [arcmin]	Amas	Date
AMI	Interferometre	15	10	1,5	100	En cours
SZA	Interferometre	30 ; 90	10	1	?	2005
AmiBA	Interferometre	90	2	1	100	2006
ACBAR	bolometrique	150 ; 219 ; 274 ; 345	3	< 3,9	10	En cours
Bolocam	bolometrique	143 ; 214 ; 273	1	1	10	En cours
APEX	bolometrique	150 ; 217	200	0,7	1000	2005
Olimpo	bolometrique	143 ; 217 ; 385 ; 600	300	~ 2	1000	2005
ACT	bolometrique	145 ; 215 ; 265	200	~ 1	1000	2005
SPT	bolometrique	95 ; 150 ; 219 ; 274 ; 345	4000	<1	20000	2008
Planck-HFI	bolometrique	100; 143; 217; 353; 545; 857	41253	5	10000	2007

FIG. 3.1.1 – Tableau, non exhaustif, de différents sondages SZ présents et à venir. Les caractéristiques sont données à titre indicatif et sont susceptibles de varier.

Bande	Nom	Couverture [deg ²]	FWHM	Date	Sources	Suivi redshifts	Redshift maximum
Optique	CFHTLS « wide »	170	~ 1"	En cours	100	100%	1,2
SZ	APEX	200	0,7'	2005	1000	?	2
	Olimpo	300	< 3,5'	2005	1000	Photo	2
	ACT	200	~ 0,9'	2005	1000	?	?
	SPT	4000	< 1'	2008	20000	?	?
	Planck-HFI	41253	5'	2007	10000	?	?
X	XMM-LSS	5,5	6"	En cours	120	50%	1,2
	Chandra	?	qqc arcsec	En cours	?	?	?
	ROSAT-REFLEX	41253	qqc arcsec	Terminé	452	100%	0,3
Weak-shear	CFHTLS « wide »	170	~ 1"	En cours	< 50 ?	100%	0,9
	DUNE	20000	~ 0,23"	2011	?	Photo au sol	1,2
Galaxies	2df	1500	qqc arcsec	En cours	250000	100%	0,2
	SDSS	?	?	En cours	?	100%	?

FIG. 3.1.2 – Récapitulatif des sondage d'amas grand champ en optique, X, SZ et weak-shear.

3.2 L'instrument Olimpo

Olimpo est une collaboration Européenne constituée en grande majorité de laboratoires Italiens. La collaboration s'étend au laboratoire Anglais de Cardiff et aux laboratoires français : Dapnia/SPP au CEA et CSNSM d'Orsay. On peut trouver des informations sur l'expérience dans Masi et al. (2003b) et Masi et al. (2003a). L'expérience Olimpo a 3 objectifs scientifiques majeurs :

- la détermination du spectre primordial de puissance des anisotropies du CMB.
- l'observation des amas de galaxies par effet SZ : deux types d'observations seront réalisées. Tout d'abord des observations pointées vers les amas proches afin de sonder leur structure interne, ensuite une observation "grand champ" afin de cartographier les amas lointains.
- l'observation des galaxies infrarouges : ce sont des sources d'avant-plan relativement lointaines ($z < 1$), émettant dans le domaine millimétrique.

C'est le domaine des observations SZ "grand champ" qui m'intéresse et qui est exposé ici. Nous espérons qu'Olimpo effectuera son premier vol de test en décembre 2005 avec une couverture du ciel de l'ordre de 25 deg^2 , le vol scientifique étant prévu courant 2006 avec une couverture de 300 deg^2 . L'instrument Olimpo est en grande partie extrapolé de l'expérience *Boomerang*. C'est une expérience embarquée en ballon stratosphérique, optimisée pour un vol longue durée de 10 jours. La figure 3.2.1 montre la nacelle, le système optique chaud et le cryostat.

3.2.1 Telescope

La nacelle de 1.5 tonnes est emportée par un ballon d'hélium à 40 km d'altitude. Le système optique est constitué d'un miroir primaire de 2.6 m de diamètre permettant une résolution spatiale de $3'$ à 143 GHz. Le miroir secondaire est porté par un réseau de cables en kevlar de faible section. La partie froide de l'optique comprend un miroir Lyot stop permettant d'uniformiser le faisceau avant son entrée dans les détecteurs et de limiter les effets de bords de l'optique chaude.

3.2.2 Cryostat

Le cryostat utilisé est celui de BOOMERanG qui a prouvé ses bonnes performances pour un vol de longue durée à la température de 0.3 K (Masi et al., 1999). Le volume disponible pour le système optique froid et les détecteurs est de 50 l. La figure 3.2.2 montre le système optique froid et l'intérieur du cryostat.

3.2.3 Bolomètres

Les détecteurs sont placés au foyer du système optique. Chacun est une matrice de bolomètres refroidie à 300 mK. Ils sont conçus pour l'observation du ciel aux quatre fréquences optiques suivantes : 143, 217, 385 et 600 GHz. Pour donner un ordre de grandeur, Olimpo sera sensible à des variations de température de l'ordre de $15 \mu\text{K}$ à 143 GHz dans le lobe d'observation. On espère que la bande passante optique soit de type "top hat", de largeur 30 GHz indépendante de la fréquence. Cette largeur de

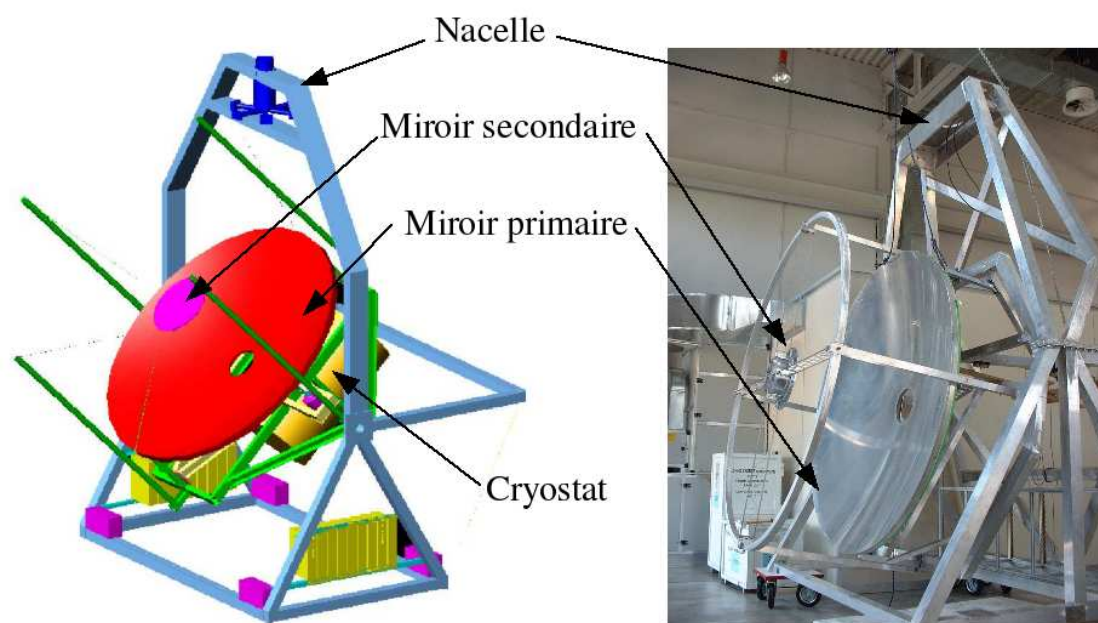


FIG. 3.2.1 – Schéma et photo d'Olimpo. On distingue sur le schéma : l'optique chaude (miroir primaire et secondaire), la mécanique de pointage du ballon et la nacelle, et le cryostat derrière le miroir primaire. La photo montre principalement la nacelle, les miroirs primaire et secondaire portés par les bras en kevlar de faible section.

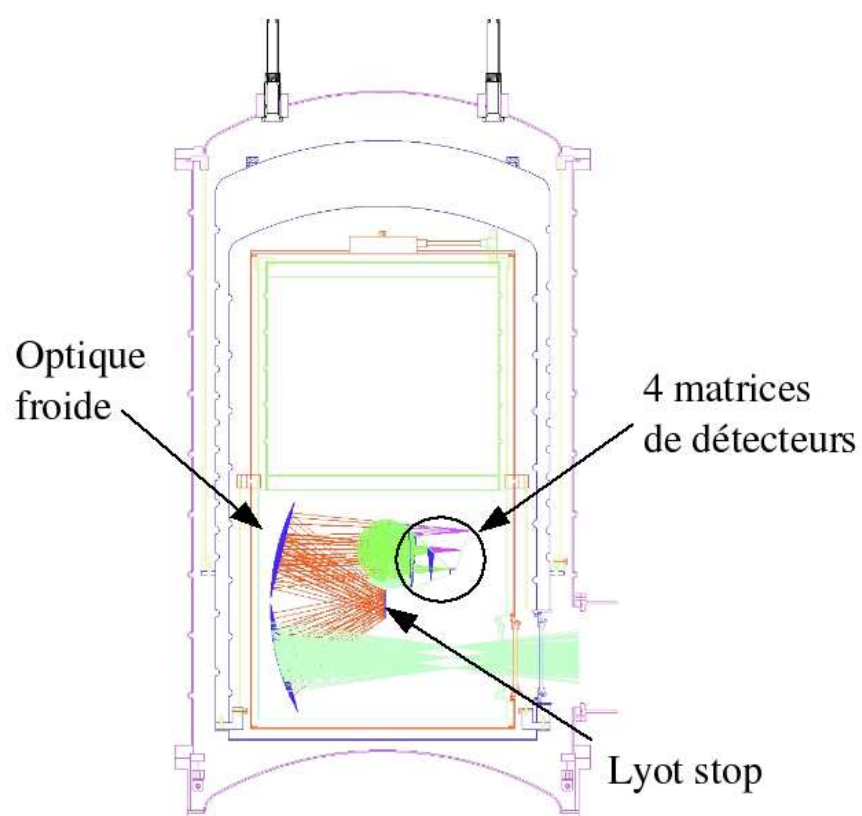


FIG. 3.2.2 – Schéma du cryostat d'Olimpo, adapté de celui de BOOMERanG. L'optique froide y est figurée.

Bande ν [GHz]	143	217	385	600
Largeur de bande $\Delta\nu$	30	30	30	30
N bolomètres	19	19	19	19
FWHM [arcmin]	3	2	2	2
Noise level [$\mu\text{Ks}^{1/2}$]	150	200	500	5000

TAB. 3.1 – Caractéristiques attendues pour l'expérience Olimpo.

bande est obtenue par la présence de cônes et de filtres optiques limitant la bande passante de l'optique. Les détecteurs sont éclairés par deux divisions successives du front d'onde grâce à des lames semi-réfléchissantes, ceci est visible sur le schéma 3.2.2.

Le tableau 3.1 résume les caractéristiques d'Olimpo.

3.2.4 Système de pointage

Le système de pointage (ACS pour *Attitude Control System*) est basé sur un scan azimuthal rapide du ciel (comme dans BOOMERanG). Ce système à deux rôles :

- diriger le pointage grâce à des moteurs de précision,
- mesurer précisément cette direction afin de permettre une reconstruction du pointage une fois le vol terminé.

Les instruments permettant de reconstruire ce pointage sont : un système de GPS différentiel et un gyroscope. Une camera CCD embarquée fournira les pointages d'étoiles brillantes dans le champ et dont les positions sont bien connues pour l'étalonnage du système. Avec ce système la précision de pointage attendue est meilleure que la minute d'arc.

3.3 Comparaison avec APEX et Planck

Les expériences Olimpo, APEX et Planck, ont les mêmes objectifs scientifiques et, bien que n'ayant pas les mêmes stratégies d'observation, utilisent des matrices de bolomètres. Ces trois expériences seront opérationnelles entre 2006 et 2008. La figure 3.3.1 montre les deux expériences APEX et Planck.



FIG. 3.3.1 – Photo et schéma des expériences APEX (gauche) et Planck (droite).

- APEX est une expérience au sol, située au nord du Chili sur le plateau d'Atacama. Elle met en oeuvre une antenne de 12 m de diamètre. Cette antenne permet d'atteindre une résolution de $0.7''$. Le bruit atmosphérique limite les observations aux fréquences 143 et 217 GHz. A ces fréquences, l'encombrement n'étant pas un facteur limitant, la caméra ASZCA (APEX-SZ) développée à Berkeley, comportera 300 bolomètres répartis entre les deux fréquences. Le choix des fréquences est encore en discussion. La résolution spatiale très bonne limite la taille du sondage et on peut espérer en deux ans observer 170 deg^2 du ciel.
- Olimpo suit une autre approche. C'est une expérience ballon qui observera le ciel dans quatre bandes de fréquences (143, 217, 385 et 600 GHz). Le télescope de 2.6 m de diamètre limite la résolution angulaire à deux minutes d'arc pour les 3 fréquences les plus hautes et $3.5''$ pour la fréquence la plus basse. Olimpo observera 300 deg^2 du ciel en 10 jours.
- Planck est une expérience satellite. L'expérience, non contaminée par l'atmosphère, observera le ciel dans dix bandes de fréquence couvrant la gamme 90 – 800 GHz avec une sensibilité excellente. La taille de l'antenne (moins d'un mètre) limite la résolution angulaire à cinq minutes d'arc au mieux. Planck observera, en un an, l'ensemble du ciel conduisant au plus grand catalogue

d'amas SZ disponible.

Ces trois expériences, aux caractéristiques différentes permettront des observations SZ complémentaires : d'un sondage plein ciel à des observations pointées d'amas à haute résolution.

Le tableau de la figure 3.3.2 présente les caractéristiques utiles aux observations SZ de ces trois expériences. La simulation présentée dans la seconde partie de ce

Expérience	Fréquences	Sensibilité [$\mu\text{K s}^{1/2}$]	FWHM	N bolo.	Couverture	Durée
Olimpo	143	150	3"	19	300 deg ²	10 jours
	217	200	2"	35		
	385	500	2"	35		
	600	5000	2"	35		
APEX	150	280	0.7"	162	170 deg ²	2 ans
	217	280	0.7"	162		
Planck	143	70	7.1"	12	43000 deg ²	1 an
	217	100	5"	12		
	353	300	5"	12		
	545	2200	5"	12		
	Autres fréquences					

FIG. 3.3.2 – Tableau illustrant les différentes caractéristiques instrumentales des prochaines observations d'amas SZ grand champ, Olimpo, APEX et Planck.

manuscrit (chapitre 4) permet de simuler des observations types de ces trois instruments. Un résultat préliminaire, présenté en conclusion de ce manuscrit, montre que dans l'expérience APEX, l'absence de haute fréquence obligera à développer des algorithmes spécifiques pour distinguer les sources ponctuelles infrarouges des amas SZ. Dans le cas de Planck et d'Olimpo, comme je le présente dans le chapitre 5, les algorithmes de séparation en composantes permettent d'éliminer ces sources ponctuelles contaminantes.

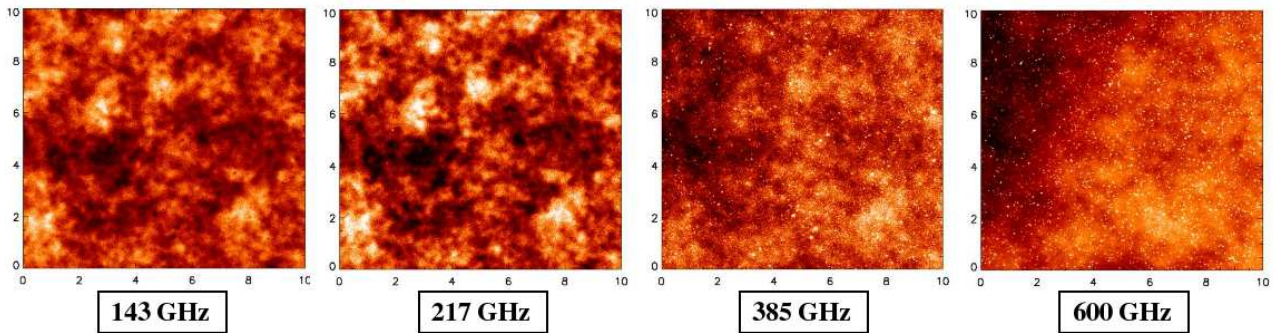
Deuxième partie

Les sondages SZ : effets de sélection et cosmologie

Dans les prochaines années, une nouvelle génération d'instruments dédiés, basés sur des grandes caméras de bolomètres (de plusieurs dizaines à plusieurs centaines de détecteurs) fourniront des observations SZ permettant la détection de plusieurs centaines d'amas de galaxies lointains. Parmi ces expériences on peut citer : APEX¹, ACT², BOLOCAM³, OLIMPO (Masi et al., 2003) qui observeront quelques centaines de degrés carrés. Par ailleurs le satellite Planck (Lamarre et al., 2004), lancé en 2007, fournira un catalogue d'amas plein ciel jusqu'à des redshifts plus grands que 1. De plus la nouvelle génération d'interferomètres sera aussi en activité avec les expériences AMI⁴, AMiBA⁵ et SZA⁶.

L'intérêt des observations SZ pour étudier la cosmologie provient des propriétés de l'effet SZ. La constance de la brillance de surface avec la distance, rendant cet effet indépendant du redshift de l'amas observé, et la mesure directe de l'énergie thermique de l'amas, devrait permettre de mesurer la fonction de masse des amas lointains. Par ailleurs des observations combinées SZ, X et par cisaillement gravitationnel permettront de limiter les erreurs systématiques, importantes pour ce type de sondage, et de comprendre statistiquement la physique interne des amas. Cette physique interne doit être bien comprise pour que la mesure des abondances d'amas de galaxies soit une sonde cosmologique compétitive, en terme de contraintes, avec les observations du CMB ou des supernovae de type Ia. La distribution des amas de galaxies en fonction du redshift est en effet sensible aux paramètres cosmologiques σ_8 et Ω_M , mais aussi au paramètre effectif de chauffage du gaz intra-amas T_* .

La figure ci-dessous (reproduction de la figure 4.5.1) présente le ciel millimétrique tel qu'attendu pour ces prochains sondages. Les cartes font 10 degrés de côté et sont



tirées de simulation du ciel millimétrique observé par Olimpo. On constate sur cette figure qu'aux deux plus basses fréquences (143 et 217 GHz) les anisotropies du CMB

¹<http://bolo.berkeley.edu/apexsz>

²<http://www.hep.upenn.edu/~angelica/act/act.html>

³http://astro.caltech.edu/~lgg/bolocam_front.htm

⁴<http://www.mrao.cam.ac.uk/telescopes/ami/index.html>

⁵<http://www.asiaa.sinica.edu.tw/amiba>

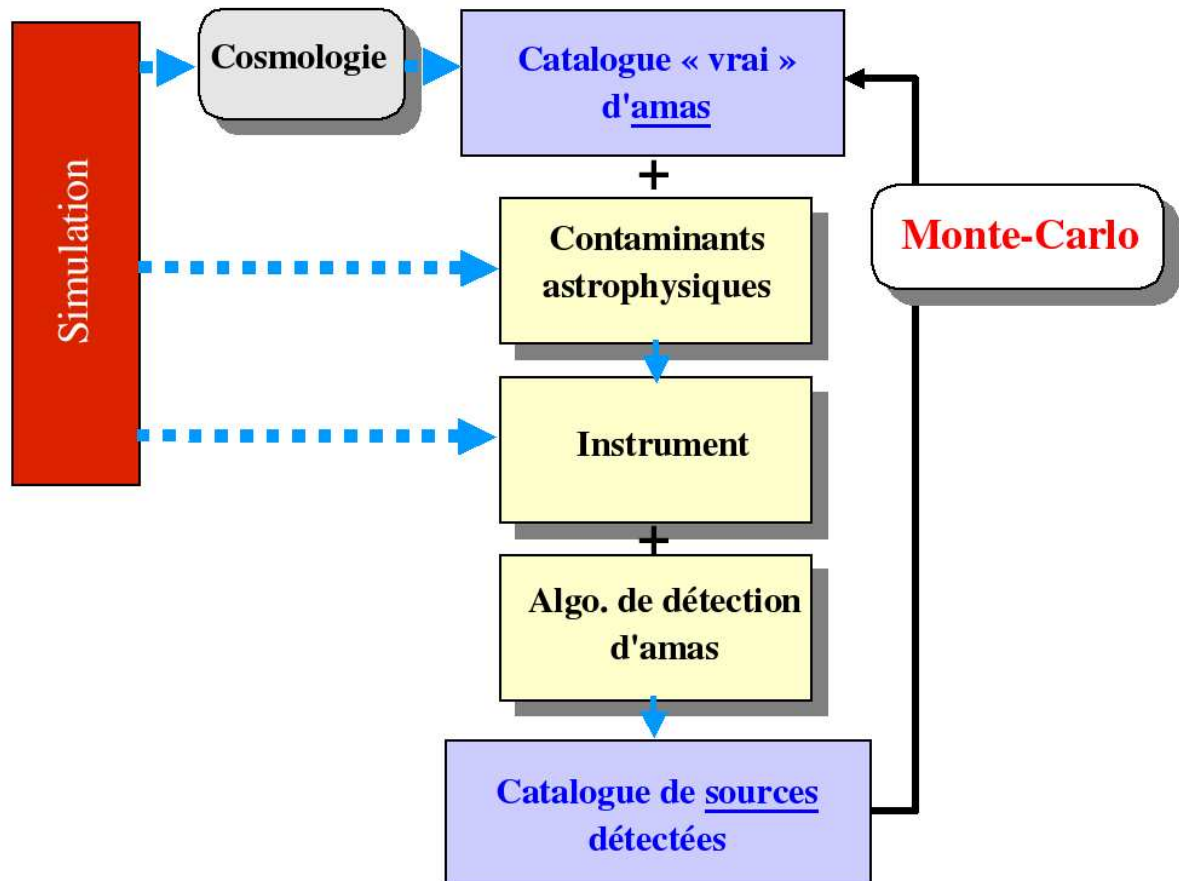
⁶<http://astro.uchicago.edu/sze>

dominent l'émission. Aux plus hautes fréquences (385 et 600 GHz) ce sont les avant plans et les sources ponctuelles qui dominent l'émission. Qu'en est-il des amas SZ ? L'observation de ces quatre cartes suggère deux points importants :

1. tout d'abord qu'il est nécessaire de développer des algorithmes dédiés pour détecter les amas de galaxies par effet SZ,
2. ensuite qu'on doit se poser la question des amas qui seront effectivement détectables dans de telles conditions d'observation ?

Répondre à cette question c'est étudier les effets de sélection des sondages SZ. Melin et al. (2005) a présenté une première étude des effets de sélection des observations SZ en grand champ qui se révèlent complexes.

Dans cette seconde partie je présente la chaîne d'outils que nous avons mis en place pour étudier ces effets de sélection, et notamment leur impact sur l'estimation des paramètres cosmologiques. Le schéma suivant présente les différents éléments de cette chaîne. La démarche est la suivante :



1. par simulation on se place dans le cadre d'un **modèle cosmologique**,
2. via les modèles de formation de structure on peut prédire un **catalogue vrai d'amas** dans l'angle solide de l'observation,

3. l'observation du ciel millimétrique fait intervenir des **contaminants astrophysiques** introduits dans la simulation,
4. le ciel millimétrique est observé par un **instrument** possédant ses propres caractéristiques et imperfections qui sont simulées,
5. enfin les amas sont isolés grâce à des **algorithmes de détection** qui présentent eux aussi des imperfections.

Le **catalogue de sources détectées** est comparé dans une boucle Monte-Carlo au catalogue d'amas vrai issu de la cosmologie. Les effets de sélection quantifient les “déformations” existantes entre les deux catalogues.

Mon travail a consisté en la mise au point de cette boucle Monte-Carlo, en la construction d'un modèle pour les effets de sélection, et enfin, en l'utilisation de ce modèle pour étudier l'impact des effets de sélection sur l'estimation des paramètres cosmologiques.

Dans les chapitres suivants je présente la simulation de ciel millimétrique développée au laboratoire (**chapitre 4**) permettant de prendre en compte les contaminants astrophysiques principaux et les effets instrumentaux d'observation. Dans le chapitre suivant (**chapitre 5**) je présente la méthode de détection d'amas de galaxies développée en collaboration avec le service d'astrophysique SAp du Dapnia (Pires et al., 2005). Une fois les éléments de la boucle Monte-Carlo introduits, le chapitre suivant (**chapitre 6**) présente les effets de sélection de ces sondages et le modèle qui en est fait, tandis que le dernier chapitre (**chapitre 7**) montre l'utilisation de ce modèle d'observation comme estimateur des paramètres cosmologiques (Juin et al., 2005).

Chapitre 4

Simulation d'observation

Sommaire

4.1	Le ciel sub-millimétrique	76
4.2	Les cartes d'effet SZ thermique	78
4.2.1	Description des modèles utilisés	78
4.2.2	Calcul des distributions théoriques simulées	80
4.2.3	Calcul de la carte	83
4.3	Les phénomènes physiques “contaminants”	86
4.3.1	Le CMB	86
4.3.2	Les galaxies infrarouges	86
4.3.3	La poussière galactique	88
4.3.4	L'atmosphère	89
4.4	Spectres d'émissions des sources astrophysiques	90
4.5	Observations multi-fréquences bolométriques	93
4.5.1	Les fréquences d'observation	93
4.5.2	Bruit de mesure	93
4.5.3	Stratégie d'observation	95
4.5.4	Lobe d'antenne et précision de pointage	95
4.6	Conclusion	97

4.1 Le ciel sub-millimétrique

LA simulation présentée ici a plusieurs objectifs :

1. Reproduire les cartes du ciel observées aux fréquences des prochaines expériences d'observation du CMB : Olimpo, APEX, Planck. Ces expériences ayant toutes pour objectif scientifique l'observation de l'effet SZ elles incluent presque toutes les fréquences suivantes : 143, 217, 385 et 600 GHz adaptées à son observation. Nous nous intéresserons plus particulièrement aux émissivités des diverses composantes émettant à ces fréquences.
2. Inclure les principaux effets instrumentaux d'observation : bruit de mesure, nombre de détecteurs, lobe d'antenne, largeur, nombre et fréquence centrale des diverses bandes optiques, etc.

Le critère qui a orienté nos choix d'algorithmes de simulation est la rapidité de calcul. Nous voulons pouvoir rapidement simuler une observation typique pour un instrument afin de calculer sa fonction de sélection dans une boucle Monte-Carlo.

Dans la suite du chapitre je présente d'abord, les différentes méthodes et modèles que nous avons utilisés pour simuler les différents phénomènes astrophysiques. Ensuite je rassemble les informations concernant les différents spectres d'émissivité de ces composantes qui permettent d'aboutir aux cartes du ciel qu'observera l'instrument. Enfin je présente le modèle d'instrument que nous avons utilisé. Les figures 4.1.1 et 4.5.1 sont le résultat d'une simulation d'observation typique de l'expérience Olimpo obtenue en moins de deux minutes.

Dans la bande du spectre comprise entre 100 et 600 GHz les composantes les plus brillantes du ciel sont :

- le CMB provenant de la couche de dernière diffusion (voir chapitre 1),
- les galaxies à travers leur rayonnement infrarouge (IR) et submillimétrique,
- la poussière galactique : la poussière chaude de la Voie-Lactée émet dans les longueurs d'ondes millimétrique. Olimpo pointera hors du plan galactique où ce rayonnement est important à ces fréquences.
- Parmi ces différentes émissions les amas de galaxies sont observables grâce à l'effet SZ (voir chapitre 2).

Nous avons simulé des cartes de ces différentes émissions, observables dans le domaine millimétrique, dans le cadre de modèles cosmologiques et de formation de structures donnés. Les cartes simulées peuvent s'écrire sous la forme :

$$S_{phys}(\vec{x}; \vec{P})$$

avec

$$phys = CMB, SZ, IR, Dust$$

où \vec{x} est la direction observée dans le ciel et \vec{P} l'ensemble des paramètres des modèles mis en jeu. Pour chacune des composantes simulées j'introduirai les observations qui ont guidé le choix du modèle utilisé, la physique mise en jeu et j'explicitrai le modèle lui-même.

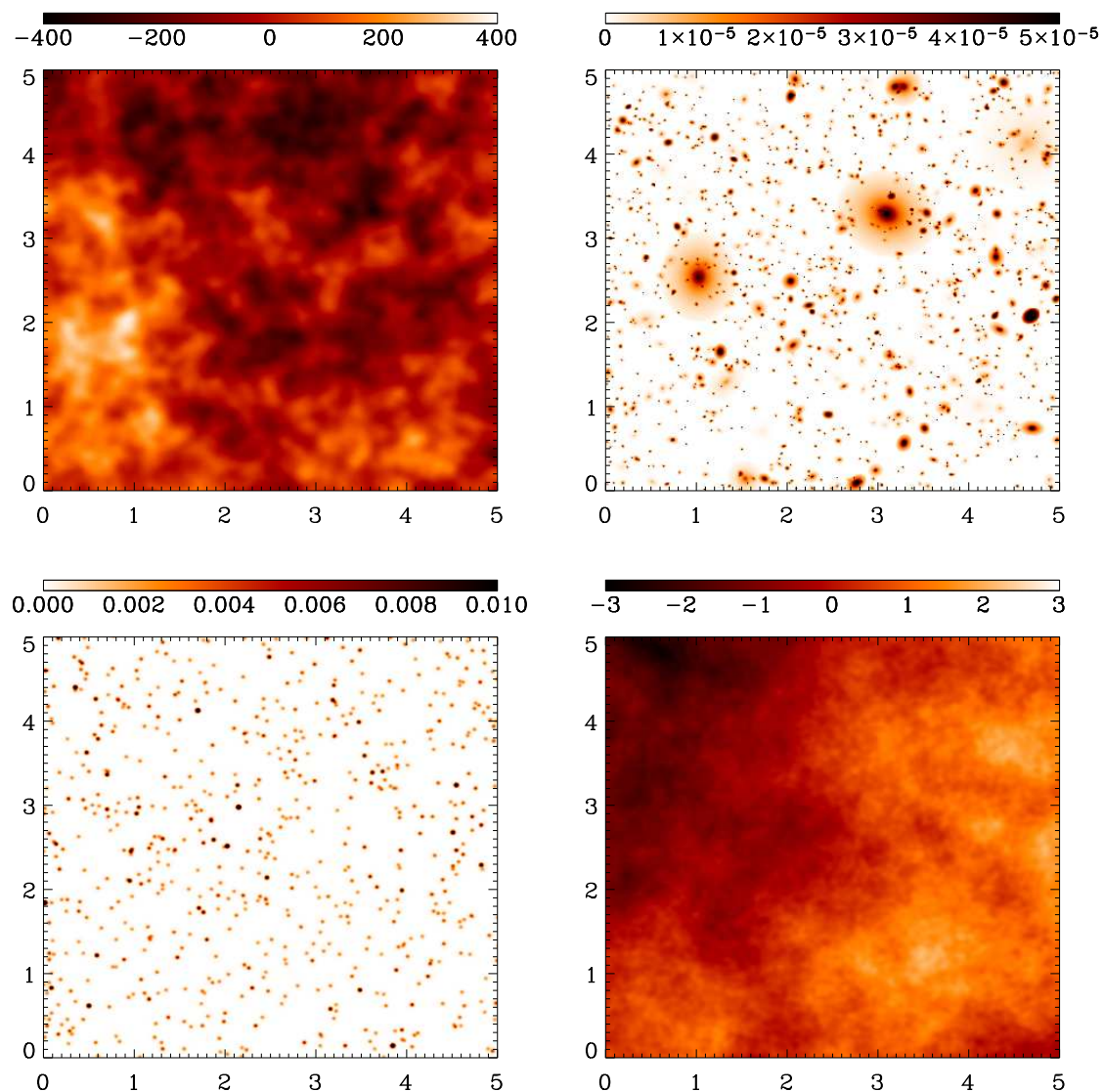


FIG. 4.1.1 – Les quatre phénomènes physiques principaux simulés émettant entre 100 et 600 GHz. De haut en bas et de gauche à droite : le CMB en μK , l'effet SZ sans unité (paramètre de Compton), les galaxies infrarouges en Jy à 350 GHz et la poussière galactique en MJ/Sr à $100\ \mu\text{m}$. Les abscisses et ordonnées sont données en degrés.

4.2 Les cartes d'effet SZ thermique

La simulation de cartes d'effet SZ passe dans un premier temps par la simulation des processus de formation des amas. Comme nous l'avons décrit dans le chapitre 2 il existe deux approches pour simuler la formation de structures massives : les simulations à N-corps (Jenkins et al., 2001), (Evrard et al., 2002) ou les simulations semi-analytiques (Bouchet et al., 1995), (Aghanim et al., 2001), (Kneissl et al., 2001). Le travail de thèse de J.-B. Melin a conduit à la mise en place d'une simulation semi-analytique sur laquelle je reviendrai au cours de ce chapitre. Elle est décrite dans Melin et al. (2005) et Melin (2004). Les premières ont l'avantage d'être plus réalistes et reproduisent correctement les corrélations entre structures. Cependant elles sont extrêmement coûteuses en temps de calcul et l'introduction de la physique baryonique dans les simulations "hydrodynamiques" est encore difficile. Les secondes forment la base de la méthode que nous choisissons ici. Elles permettent d'obtenir rapidement les fonctions de masse des structures et, moyennant des modèles de la physique tardive décrivant l'état d'équilibre des amas, elles permettent de produire des cartes de la matière sombre ou de la matière baryonique. L'inconvénient est que les effets de corrélation entre structures ne sont pas naturellement présents.

Pour atteindre notre objectif qui est de simuler la fonction de sélection d'un instrument et de la chaîne de détection d'amas (chapitres 6 et 5) utilisée, il faut pouvoir effectuer rapidement ces simulations de cartes d'amas. Cela nous conduit à préférer une méthode semi-analytique basée sur l'utilisation des différents modèles décrits ci-dessus. Tous les programmes sont écrits en C++, permettant des calculs en parallèle sur n'importe quelle ferme de PC.

La simulation des cartes SZ suit les étapes suivantes :

1. choix de modèles : cosmologique, de formation de structures et de physique d'équilibre de l'amas,
2. calcul de la distribution théorique des amas en fonction de leurs masse et redshift, puis calcul de cette distribution en terme d'observables pertinentes : paramètre de Compton intégré et redshift,
3. tirage aléatoire d'un catalogue d'amas à partir de cette distribution théorique,
4. placement des amas sur la carte et calcul des cartes individuelles des amas à partir de leur profil de densité.

La figure 4.2.1 donne un exemple pour trois zooms de cartes SZ simulées.

4.2.1 Description des modèles utilisés

La méthode semi-analytique utilisée ici pour la simulation de ces cartes se base sur le formalisme de la fonction de masse pour prédire la distribution des amas en fonction de leur masse et de leur redshift. Ce formalisme a été décrit dans la section 2.3. Ce modèle dépend des paramètres cosmologiques de l'univers dans lequel les structures évoluent. La structure interne des amas est décrite par d'autres modèles : de chauffage du gaz, d'équilibre thermique, etc. (voir le chapitre 2).

Les cartes présentées dans ce chapitre sont calculées à partir du modèle cosmologique et de formation de structures \vec{P} présenté dans le tableau 4.1.

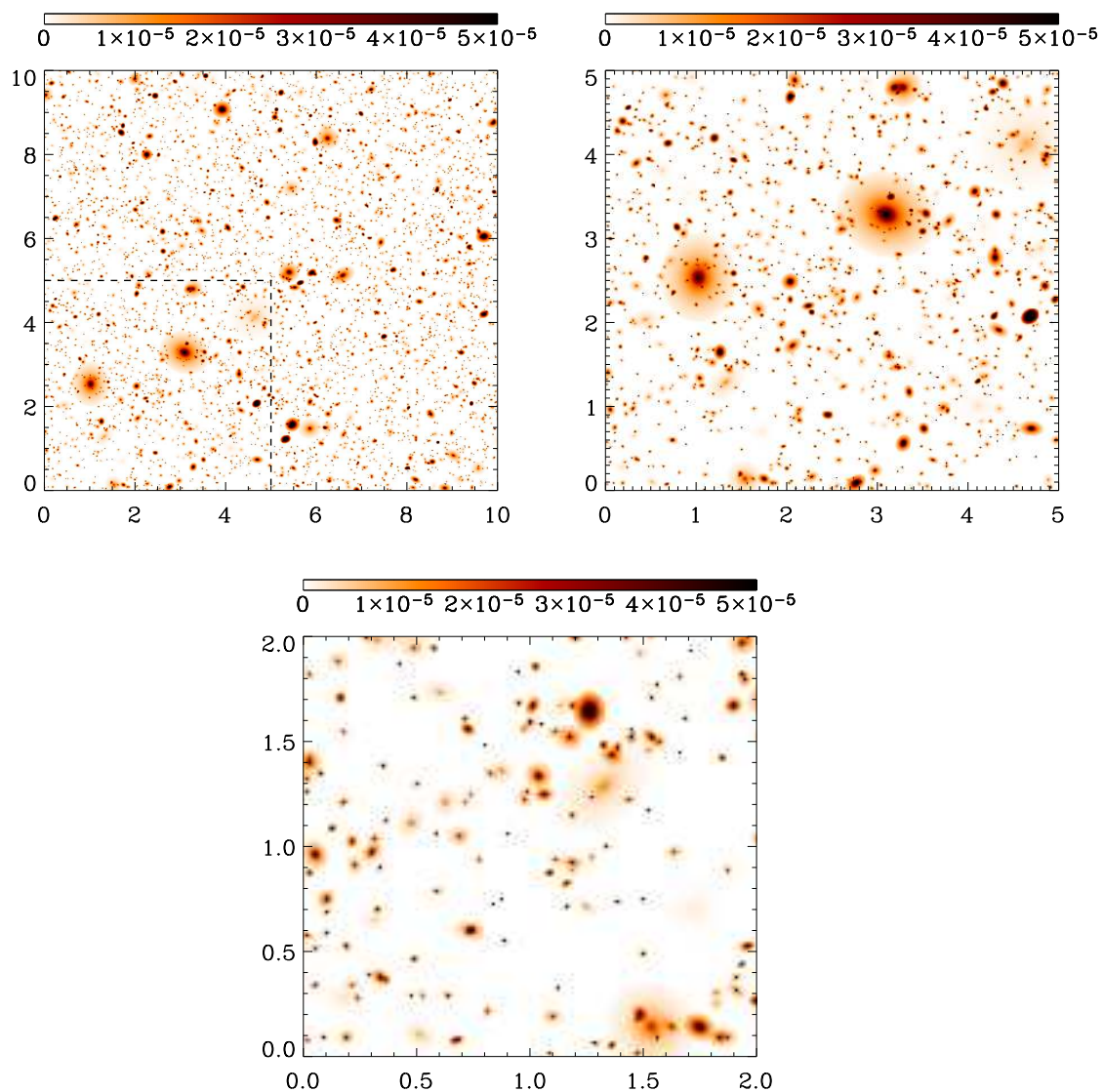


FIG. 4.2.1 – Exemple de cartes simulées d'amas de galaxies observés par effet SZ. Les grandeurs représentées sur la carte sont sans unité. La grandeur donnée est le paramètre de Compton y sur la ligne de visée. Les abscisses et ordonnées sont en degrés. La résolution de ces cartes est de 6400 pixels/deg².

Ω_{tot}	Ω_b	Ω_Λ	Ω_{DM}	h	n_s	σ_8	f_{masse}	T_\star	f_g
1	0.04	0.7	0.26	0.7	1	1.0	S. & T.	1.6	$0.9 \Omega_b / \Omega_m$

TAB. 4.1 – Cosmologie utilisée pour la génération des cartes présentées dans cette section et dans le chapitre 5. Ω_{tot} est la densité totale de l'univers, Ω_b est la densité de baryons, Ω_Λ est la densité du vide, Ω_{DM} est la densité de matière noire, h est la constante réduite de Hubble, n_s est l'indice spectral du spectre primordial de puissance, σ_8 est la normalisation de ce spectre calculée sur une échelle de 8 Mpc, f_{masse} est la fonction de masse utilisée pour le calcul de l'abondance des amas, T_\star est la normalisation de la relation masse-température dans les amas, et f_g est la fraction de gaz dans les amas. Ces différents paramètres sont présentés dans les chapitres 1 et 2.

Notre simulation calcule l'effet SZ thermique car la sensibilité des prochaines expériences ne permet pas de mesurer les autres effets SZ dans les sondages grand champ (voir 2.4). Pour cela nous devons nous donner des modèles de physique décrivant l'état d'équilibre des amas afin d'en déduire leur structure interne. Nous utilisons les modèles décrits dans le chapitre 2. Les satellites XMM et Chandra ont montré que la structure des amas est complexe, leur forme est le plus souvent elliptique, et qu'ils comportent des sous-structures. Le coeur est généralement plus froid que la partie externe (voir l'introduction du chapitre 2). Notre simulation suppose les amas elliptiques et isothermes. La notion d'ellipticité intervient dans l'utilisation de la fonction de masse de Sheth et Tormen et dans l'introduction du profil de densité des amas. Un tel modèle ne prend pas en compte les sous-structures présentes dans les amas. Cependant celles-ci seront difficilement observables sur les amas lointains (ceux qui nous intéressent dans les sondages grand champ) avec les résolutions angulaires des futures expériences, qui est de l'ordre de la minute d'arc. Par ailleurs l'hypothèse d'amas isothermes est bien vérifiée pour la région entre le rayon de coeur et le rayon viriel, zone où l'effet SZ est significatif et où il prend tout son intérêt, car complémentaire des observations X peu sensibles dans ces régions de faible densité électronique. Ces deux hypothèses simplificatrices restent intéressantes car, dans le cadre d'études statistiques, elles permettent de prédire des lois d'échelles vérifiées, en moyenne, sur une large gamme d'amas.

4.2.2 Calcul des distributions théoriques simulées

On se place dans le cadre cosmologique donné par le tableau 4.1. En utilisant la fonction de masse de Sheth & Tormen décrite dans la section 2.3 qui donne la distribution des amas de galaxies en fonction de leur masse M et de leur *redshift* z par :

$$\frac{dN}{dz dM d\Omega}(z, M; \vec{P}) = f_{masse}(z, M; \vec{P}) \frac{\rho_0 c D_A^2 (1+z)^2}{MH(z)}$$

avec :

$$f_{masse}(z, M; \vec{P}) = f_{S\&T}(z, M; \vec{P}) = A \sqrt{2a/\pi} [1 + (1/a\nu_c^2)^p] \nu_c \left| \frac{d \ln \sigma}{d \ln M} \right| e^{-\frac{1}{2} a \nu_c^2}$$

où $A = 0,3222$, $a = 0,707$ et $p = 0,3$. $\nu_c = \delta_c/\sigma(z, M)$ avec δ_c le contraste linéaire de densité nécessaire à l'effondrement de la masse M . $\sigma(z, M)$ est l'écart-type au redshift z de la distribution des perturbations de densité pour l'échelle de masse M . Son calcul est donné dans la section 2.2.1 du chapitre 2. On peut calculer cette distribution en fonction des observables SZ, flux de Compton intégré Y et taille apparente θ de la structure :

$$\frac{dN}{d\theta dY d\Omega}(\theta, Y; \vec{P})$$

On verra au chapitre 6 que l'observable "taille apparente" θ est peu exploitable car difficilement mesurable. Par ailleurs cette quantité est difficile à relier à une grandeur physique intrinsèque aux amas. Dans notre simulation elle devrait correspondre au rayon viriel apparent, mais d'une part les zones externes des amas, de faible brillance, sont noyées dans le bruit et donc estimer cette taille et son lien réel avec le rayon viriel est une tâche extrêmement difficile, et d'autre part le rayon viriel est une grandeur mal comprise de la physique des amas qui, de surcroît, dépend du choix du contraste de densité linéaire Δ_c . Parvenir à reconstruire la taille des amas constitue un travail de photométrie complexe que je n'aborde pas dans cette thèse (des pistes sont données dans le chapitre 6). On calculera donc en pratique la distribution simulée des amas en redshift z et flux Compton Y :

$$\frac{dN}{dz dY d\Omega}(z, Y; \vec{P})$$

C'est un choix que nous faisons ici de ne pas tenir compte de la taille apparente des amas mais plutôt de choisir comme grandeurs observables le flux des amas et leur redshift. Cette approche nécessitera, pour les prochaines expériences SZ d'avoir un suivi permettant de connaître le redshift des amas détectés. L'impact de ces mesures de redshift complémentaires au sondage SZ est étudiée au chapitre 7. En pratique on calcule le nombre d'amas dans une gamme de redshift donnée et ayant un paramètre de Compton minimum :

$$N = \int_{Y > Y_{min}} \int_0^\infty \int_\Omega \frac{dN}{dz dY d\Omega}(z, Y; \vec{P}) dY dz$$

La valeur de Y_{min} dépend des caractéristiques de l'instrument considéré. Je reviendrai sur ces valeurs au chapitre 6. À partir de cette distribution théorique on tire, selon une statistique de Poisson, un nombre d'amas observés dans l'angle solide du sondage :

$$N_o = \text{Poisson}(N) * \Omega_{\text{sondage}}$$

On génère ensuite l'ensemble du catalogue en tirant les quantités z et Y selon la distribution théorique précédente par une méthode simple d'acceptance-rejection à deux dimensions. La figure 4.2.2 montre un exemple de catalogue produit de cette manière. Ce catalogue contient les différentes informations physiques : paramètre de Compton intégré Y , rayon de coeur R_c des amas, température électronique T_e , masse M , redshift z , rayon viriel R_v .

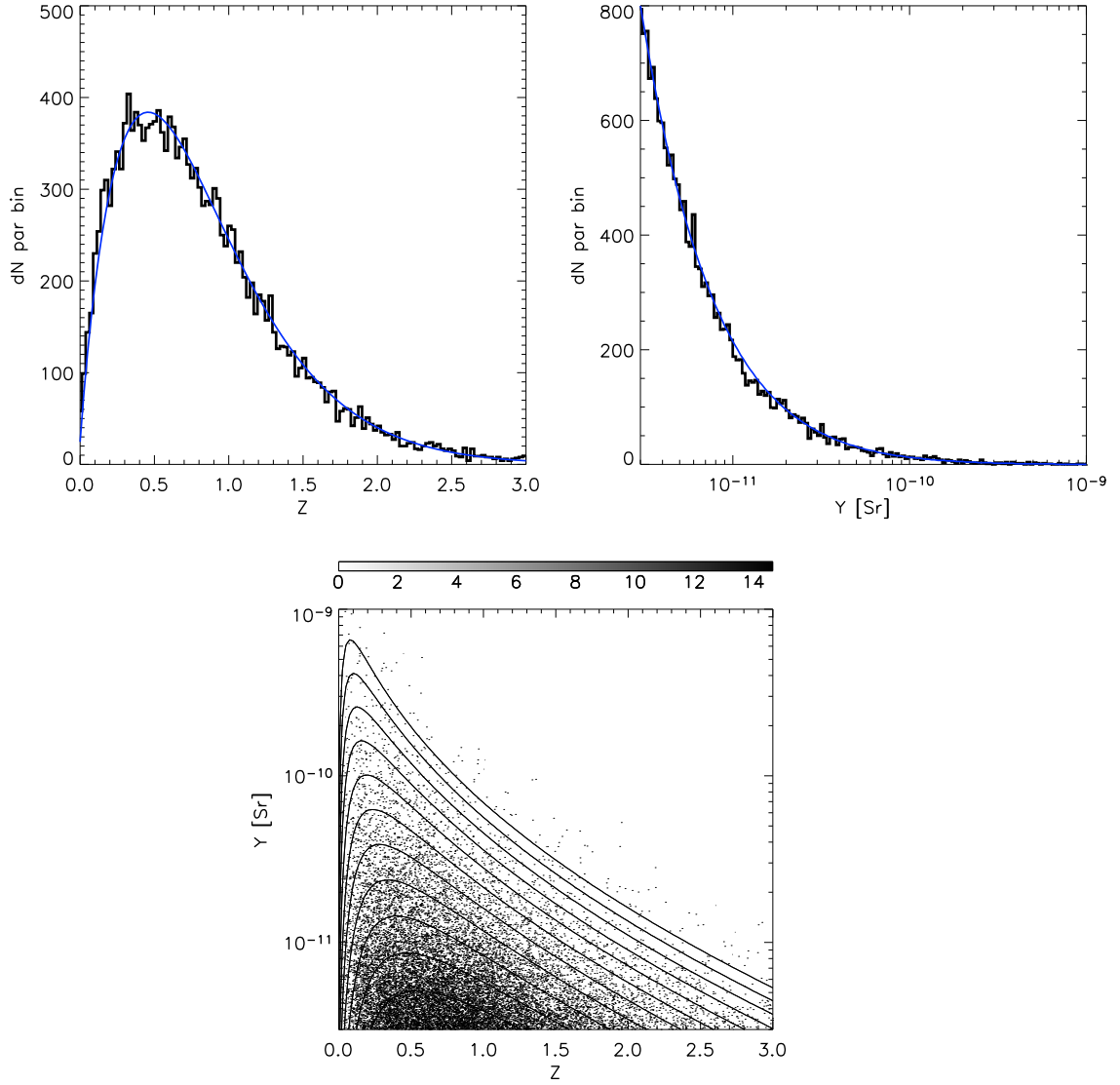


FIG. 4.2.2 – Exemple de génération d'un catalogue d'amas. Le nombre d'amas est donné par bin de redshift et de paramètre de compton intégré pour un total de 17704 amas générés sur 400 deg^2 . Les contours donnent l'abondance théorique $\frac{dN}{dzdY}(z, Y > Y_{min})$ tandis que le nuage de points représente la génération aléatoire du catalogue.

Ces grandeurs physiques sont calculées grâce aux relations présentées à la section 2.2.2 et dérivent du modèle d'amas sphérique isotherme :

$$kT_e = kT_\star \left(\frac{M}{10^{15} M_\odot h^{-1}} \right)^{1/3} (\Delta_c E^2)^{1/3} \left(1 - 2 \frac{\Omega_\Lambda}{\Delta_c} \right) \text{ keV} \quad (4.1)$$

où Δ_c est défini comme étant la sur-densité moyenne non-linéaire à l'intérieur du rayon viriel présenté à la section 2.2.1. Sa valeur dépend faiblement de la cosmologie et du redshift. Dans un univers de type Einstein-de Sitter son calcul analytique est possible et donne $\Delta_c = 18\pi^2$. La grandeur $E^2 = \Omega_m(1+z)^3 + \Omega_\Lambda + \Omega_k(1+z)^2$ est définie au chapitre 1.

Le paramètre de Compton intégré des amas est calculé à partir de la masse et du redshift de l'amas (voir l'équation 2.71). Précisément :

$$Y = \frac{k\sigma_T}{m_e c^2} \frac{1}{D_a^2} \frac{1+2/\mu}{5m_p} f_g \frac{\Omega_b}{\Omega_M} T_e M \quad (4.2)$$

En suivant l'approche de Ma and Fry (2000), nous choisissons dans notre simulation $r_v = r_{200}$: le rayon viriel est le rayon incluant une surdensité valant 200 fois la densité moyenne de l'univers. On calcule que :

$$R_v = \frac{1.63 \cdot 10^{-5}}{\Omega_M^{1/3}} \left(\frac{M}{h^{-1} M_\odot} \right)^{1/3} h^{-1} \text{ Mpc} \quad (4.3)$$

4.2.3 Calcul de la carte

Pour la création de cartes, le catalogue décrit ci-dessus est enrichi de nouvelles informations :

1. les positions des amas dans le champ observé,
2. les paramètres du profil beta permettant de décrire la densité électronique intra-amas,
3. les ellipticités des amas et l'orientation de leur grand axe.

Positionnement des amas. Dans la simulation que nous utilisons, semi-analytique, la position des amas sur la carte est tirée uniformément dans l'angle solide du champ observé. Nous n'introduisons pas de corrélation entre les positions des amas.

Profil beta. Les modèles précédents permettent de décrire les grandeurs physiques "intégrales" de l'amas : masse totale, paramètre de Compton intégré, température moyenne, etc. Pour construire les cartes d'effet SZ nous avons besoin du paramètre de Compton sur la ligne de visée : y . Il s'exprime de la manière suivante (voir le chapitre 2 section 2.4) :

$$y = \int_{los} \frac{kT_e}{m_e c^2} n_e(r) \sigma_T dl \quad (4.4)$$

où l'intégrale est calculée le long de la ligne de visée (los pour *line of sight*). T_e est la température électronique de l'amas, et m_e la masse de l'électron. n_e est la densité électronique de l'amas au point $r = \sqrt{l^2 + b^2}$ où b est le paramètre d'impact de la ligne de visée par rapport au centre de l'amas, tandis que σ_T est la section efficace de Thomson. Dans le cas d'amas isothermes cette équation se ramène à :

$$y = \frac{k\sigma_T}{m_e c^2} T_e \int_{los} n_e(r) dl \quad (4.5)$$

Ainsi la connaissance du profil de densité $n_e(r)$ est nécessaire à la construction des cartes SZ. Nous utilisons dans notre simulation le profil de King présenté à la section 2.2.2.1 :

$$n_e(r) \begin{cases} \propto \left(1 + \frac{r^2}{R_c^2}\right)^{-3\beta/2} & r < R_v \\ = 0 & r > R_v \end{cases} \quad (4.6)$$

Les paramètres choisis pour ce profil sont : $\beta = 2/3$ et $\alpha = 10$. On a alors pour le rayon de coeur : $R_c = \frac{R_v}{\alpha} = \frac{R_v}{10}$. La projection intégrée de ce profil est donnée par :

$$\zeta(\theta) = \left(1 + \frac{\theta^2}{\theta_c^2}\right)^{1/2-3/2\beta} \frac{J\left[\left(\frac{\alpha^2-\theta^2/\theta_c^2}{1+\theta^2/\theta_c^2}\right)^{1/2}, \beta\right]}{J(\alpha, \beta)} \quad (4.7)$$

avec $J(a, b) = \int_0^a (1+x^2)^{-3b/2} dx$ et $\theta_c = R_c/D_a$, le rayon de coeur apparent. En considérant le cas où : $\alpha = 10$ et $\beta = 2/3$ on a :

$$J(a, 2/3) = \tan^{-1}(a) \quad (4.8)$$

et

$$\zeta(\theta) = \begin{cases} \left(1 + \frac{\theta^2}{\theta_c^2}\right)^{-1/2} \frac{\tan^{-1}\left(\frac{100-\theta^2/\theta_c^2}{1+\theta^2/\theta_c^2}\right)^{1/2}}{\tan^{-1}(10)} & \theta < \theta_v \\ 0 & \theta > \theta_v \end{cases} \quad (4.9)$$

avec $\theta_v = R_v/D_a$, le rayon viriel apparent de l'amas. On utilise cette grandeur pour calculer le paramètre de Compton sur la ligne de visée de l'amas, à une distance apparente θ du centre :

$$y(\theta) = \frac{k\sigma_T}{m_e c^2} T_e \zeta(\theta) \quad (4.10)$$

La normalisation du profil est expliquée dans le paragraphe suivant.

Normalisation et apodisation. La valeur absolue du paramètre de Compton en un point du profil est contrainte en utilisant l'information sur le paramètre de Compton intégré Y . En effet, dans le cas d'amas isothermes, ce dernier est indépendant du profil mais dépend de la masse totale de l'amas (section 2.4). On fixe la normalisation du profil en comparant $Y = \int_{\Omega} y(\theta_{eff}) d\Omega$ et Y calculé directement à partir de la masse de l'amas (équation 2.71).

Par ailleurs on pourrait envisager d'apodiser le bord des amas afin de ne pas produire de rupture dans le profil entre le fond (nul) et l'amas. Cependant, il est difficile d'apodiser le profil de manière à conserver l'amplitude du pic central. À flux total fixe, élargir le profil implique de diminuer l'amplitude du pic. Étant donné les ordres de grandeurs sur les cartes observés des différentes composantes physiques et du bruit, il nous a semblé finalement inutile d'apodiser le bord des amas.

Ellipticité. Nous tenons compte de l'ellipticité en utilisant la fonction de masse de Sheth & Tormen, Pour le profil interne des amas nous utilisons les résultats de l'article de Cooray (2000). Selon l'orientation Ψ de l'amas et son ellipticité, on calcule une distance effective θ_{eff} au centre pour chaque pixel (X_{map}, Y_{map}) de la carte "individuelle" de l'amas :

$$\begin{cases} X_{eff} &= X_{map} * \cos(\Psi) + Y_{map} \sin(\Psi) \\ Y_{eff} &= Y_{map} * \cos(\Psi) - X_{map} \sin(\Psi) \end{cases}$$

$$\theta_{eff} = \sqrt{1/e X_{eff}^2 + e Y_{eff}^2}$$

C'est cette distance θ_{eff} qui intervient dans le calcul du profil projeté de King. Cette définition de l'ellipticité correspond à celle de Cooray, qui s'est appuyé sur l'observation d'une cinquantaine d'amas proches observés en X pour en déduire une distribution de leur ellipticité apparente (voir la figure 2.2.3). Nous choisissons de tirer l'ellipticité uniformément autour de $\langle e \rangle = 0,85$ avec une excursion de 0,1. L'orientation de l'axe principal de l'amas (celui selon lequel l'amas est allongé) est tirée uniformément dans l'intervalle $[0, 2\pi]$. Je reviens sur l'ellipticité mesurée des amas détectés dans le chapitre 5. La carte SZ est la somme des cartes individuelles d'amas ainsi calculées.

4.3 Les phénomènes physiques “contaminants”

4.3.1 Le CMB

Les phénomènes cosmologiques à l'origine des anisotropies du CMB ont été présentés à la fin du chapitre 1. Les fluctuations de température mesurées par COBE-DMR et dernièrement par WMAP ont permis d'estimer un spectre de puissance du CMB. Ce spectre est exprimé en terme de C_l , décomposition en harmoniques sphériques du signal spatial. Il permet de contraindre fortement les paramètres cosmologiques. De nombreux logiciels existent pour calculer les spectres de puissance dépendant des paramètres cosmologiques. Les plus utilisés sont CMBFAST (Seljak and Zaldarriaga, 1996) et CMBeasy (Doran, 2003). Les spectres de puissance calculés, décrivant des structures de taille allant de tout le ciel à une seconde d'arc, sont exprimés en fonction de “l'ordre” de la décomposition en harmoniques sphériques du signal. Les détecteurs bolométriques mesurent des variations de puissance, ainsi le pôle d'ordre 0 n'est pas mesuré. Dans notre simulation les fluctuations du pôle d'ordre 1 (dipole dû au mouvement de la terre dans le référentiel du CMB) sont soustraites.

Pour simuler des cartes d'anisotropies du CMB nous utilisons CMBeasy pour calculer un spectre de C_l en accord avec les paramètres cosmologiques \vec{P} jusqu'aux multi-pôles d'ordre 5000. Travaillant sur des cartes au maximum de 400 deg^2 nous avons utilisé l'approximation des petites cartes pour calculer à partir du spectre en harmoniques sphériques du signal, son spectre de Fourier. Cette approche est décrite dans White et al. (1999)

Une fois le spectre de Fourier $P(\vec{k})$ calculé on réalise un tirage aléatoire :

1. en tirant pour chaque pixel de la carte une valeur aléatoire selon une statistique normale,
2. en calculant dans l'espace de Fourier à deux dimensions le mode \vec{k} du pixel et en multipliant la variance du spectre de puissance pour ce mode par la valeur aléatoire précédente,
3. on multiplie la carte par le spectre $P(\vec{k})$,
4. on normalise la carte à la variance attendue,
5. enfin la carte finale est obtenue par transformée de Fourier inverse : c'est une génération aléatoire suivant le spectre de puissance $P(\vec{k})$.

La figure 4.3.1 illustre une simulation typique de carte du CMB dans ses unités naturelles.

Il est important de noter que les cartes d'anisotropies du CMB simulées par cette méthode ne possèdent pas de structures à des échelles plus grandes que celle de la moitié de la taille de la carte. En pratique on calcule des cartes deux fois plus grandes que l'on coupe ensuite pour éviter les effets de périodicité aux bords.

4.3.2 Les galaxies infrarouges

L'ensemble des sources non-résolues dans le domaine infrarouge constitue ce que l'on appelle le fond de rayonnement infrarouge ou CIB (*Cosmic Infrared Back-*

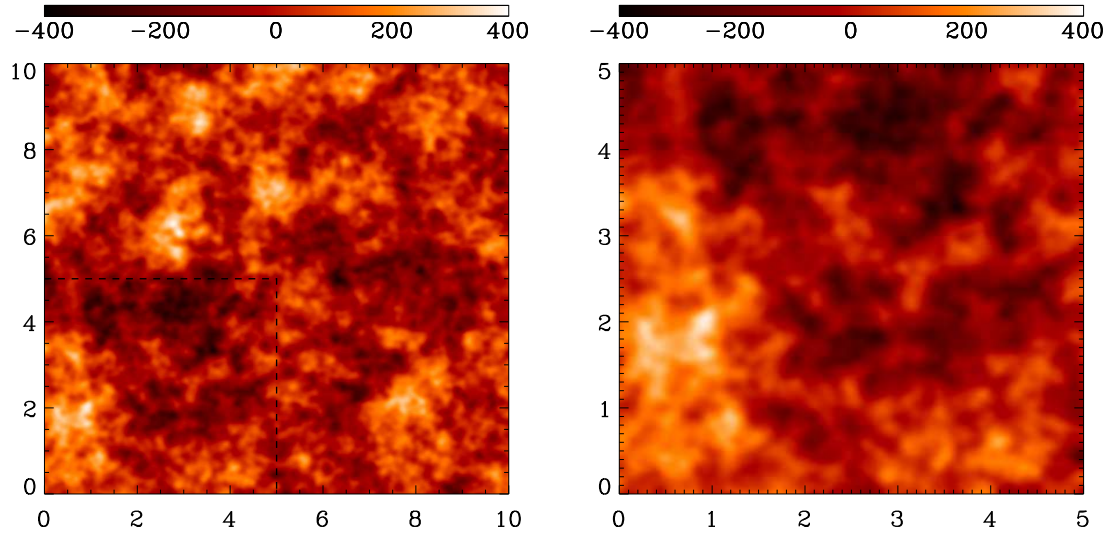


FIG. 4.3.1 – Exemple de simulation d’une carte de CMB donnée pour deux zoom différents. L’unité est le μK . Les abscisses et ordonnées sont données en degrés.

ground). Il a été détecté la première fois pas Puget et al. (1996) en analysant les cartes COBE/FIRAS. Les fluctuations du CIB ont été observées à $170\mu\text{m}$ dans le champ FIRBACK (Lagache et al., 2000). On suppose ce fond soustrait de la composante infrarouge extra-galactique que nous simulons : on simule les sources ponctuelles brillantes.

Dans la bande de fréquence sub-millimétrique on s’attend à ce que les galaxies infrarouges constituent une contamination importante pour l’effet SZ. L’expérience SCUBA a permis de constituer un catalogue de plus de 300 sources à 350 GHz (Borys et al., 2003). A partir de ces données la distribution des sources en fonction de leur flux a été modélisée sous la forme :

$$\frac{dN}{dS} = \frac{N_0}{S_0} \left[\left(\frac{S}{S_0} \right) + \left(\frac{S}{S_0} \right)^{3.3} \right]^{-1}$$

avec $S_0 = 1.8\text{ mJy}$ et $N_0 = 1.5 \cdot 10^4 \text{ deg}^{-2}$. On extrapole cette loi à des flux plus faibles que les limites d’observation de SCUBA afin d’introduire dans la simulation des sources peu brillantes.

Les galaxies sont considérées dans notre simulation comme des sources ponctuelles : elles occupent un seul pixel de la carte. Elles sont positionnées de manière uniforme dans le champ observé. On peut choisir de corrélérer la position d’une fraction d’entre elles avec les positions des amas, cependant ce travail étant encore préliminaire nous ne l’utilisons pas dans les résultats présentés ici. Un exemple de carte de galaxies infrarouges est donné dans la figure 4.3.2

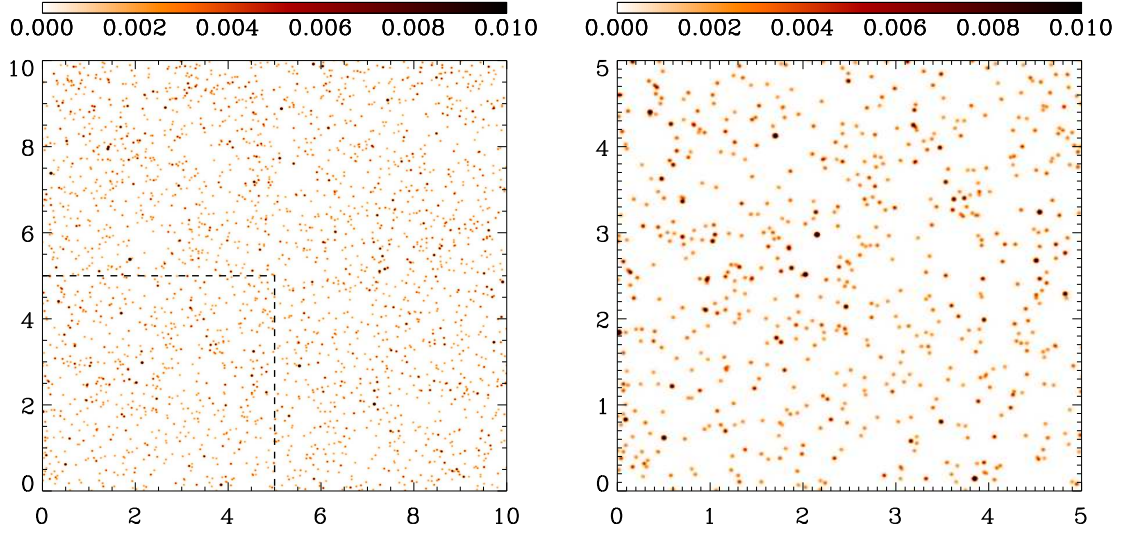


FIG. 4.3.2 – Simulation d’une carte pure de galaxies infrarouges. La carte a été lissée par un lobe de largeur à mi-hauteur de 2 minutes d’arc afin de pouvoir voir les sources ponctuelles. La carte est donnée en Jy à 350 GHz. Les abscisses et ordonnées sont données en degrés.

4.3.3 La poussière galactique

La poussière galactique est la principale source de contamination du CMB aux fréquences supérieures à 100 GHz. L’émission thermique de la poussière a été cartographiée par le satellite IRAS (*InfraRed Astronomical Satellite*), lancé en 1983, sur 96% du ciel à quatre fréquences supérieures à 3000 GHz. Par ailleurs l’instrument FIRAS (Far InfraRed Absolute Spectrophotometer) embarqué sur le satellite COBE a permis d’obtenir des mesures de l’émission de la poussière à des fréquences plus faibles voisines de celles du CMB (100 – 600 GHz). L’émission contaminante pour l’observation de l’effet SZ dans la bande de fréquence considérée est l’émission thermique de la poussière. Cette émission est due au chauffage des poussières par le rayonnement optique des étoiles. Il existe d’autres mode d’émission du milieu interstellaires : free-free, synchrotron, etc. G. Patanchon, en fait une revue dans sa thèse (Patanchon, 2003). Dans notre simulation nous incluons uniquement l’émission thermique des poussières, les autres émissions sont supposées négligeables à nos fréquences.

Pour la simulation nous voulons pouvoir générer de manière aléatoire un ensemble de cartes de poussière galactique. On suppose que les cartes de poussière galactique ont un spectre de puissance en C_l de la forme (Bouchet and Gispert, 1999) :

$$C_l \propto 1/l^3$$

Nous n’utilisons pas directement la carte de la poussière galactique (Finkbeiner and Schlegel, 1999) car nous n’aurions pas, de cette façon, une vraie génération de

cartes aléatoires. Cependant la carte obtenue par Schlegel sert à normaliser les cartes aléatoires générées par notre procédure. On normalise la moyenne des cartes générées à la moyenne d’une carte de la galaxie extraite à haute latitude galactique : 10 J/Sr observé à $\lambda = 100 \mu\text{m}$. Cette valeur dépend énormément de la latitude galactique considérée : vers le centre galactique le signal infrarouge de la poussière domine tous les autres. On utilise la même méthode que pour le CMB (4.3.1) pour calculer la carte correspondante à ce spectre de puissance. Un exemple de génération aléatoire est donné dans la figure 4.3.3.

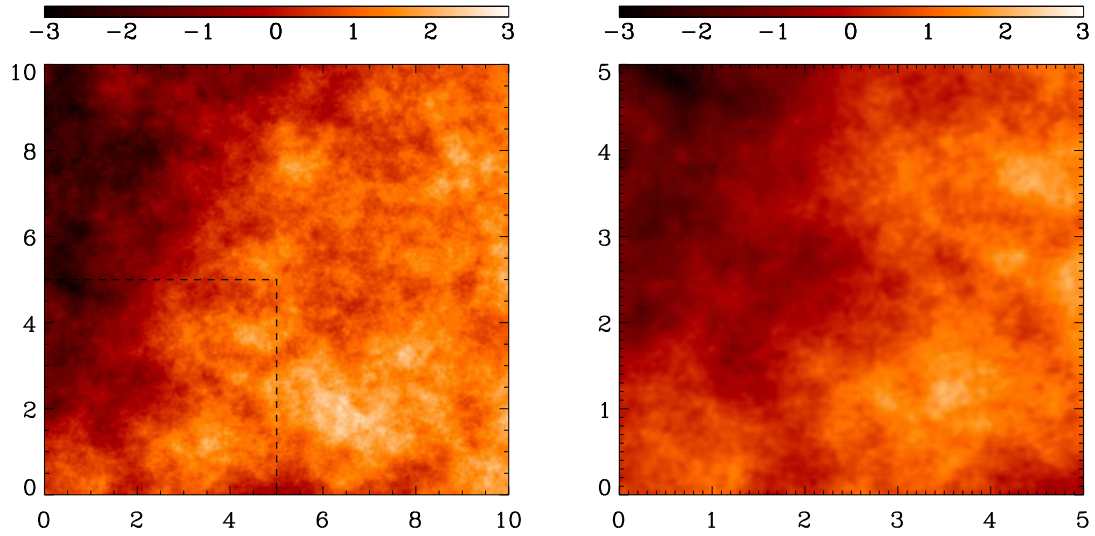


FIG. 4.3.3 – Exemple de carte simulée d’émission thermique de la poussière galactique. L’unité est le Jy/Sr à $100 \mu\text{m}$. Les abscisses et ordonnées sont données en degrés.

4.3.4 L’atmosphère

Une source de contamination importante pour les expériences bolométriques non spatiales est l’émission infrarouge de l’atmosphère. Les expériences atmosphériques sont optimisées pour réduire ses effets ou pour pouvoir les traiter facilement. Je détaille les effets de l’atmosphère et comment on peut s’en affranchir dans la section 4.5. Nous n’avons pas inclu cette émission dans notre simulation. Un modèle simple pour l’émission atmosphérique serait une dépendance proportionnelle à l’épaisseur d’air (*air-mass*) traversée pour une direction de pointage donnée. Une stratégie de pointage adaptée (à élévation constante) permet de décorréler cet avant plan des émissions astrophysiques de manière relativement simple.

4.4 Spectres d'émissions des sources astrophysiques

Dans la section précédente nous avons exposé les principales composantes physiques émettant dans la gamme de fréquences 100–600 GHz que nous avons simulées. Dans cette section je présenterai les lois d'émissivité de ces différentes composantes, ce qui nous permet de calculer les cartes aux différentes fréquences qu'observeront de futures expériences de sondage SZ (voir la figure 4.5.1) : 143, 217, 385 et 600 GHz. Les différents spectres d'émissivité sont présentés dans les paragraphes ci-dessous. Le schéma 4.4 illustre ces différents spectres.

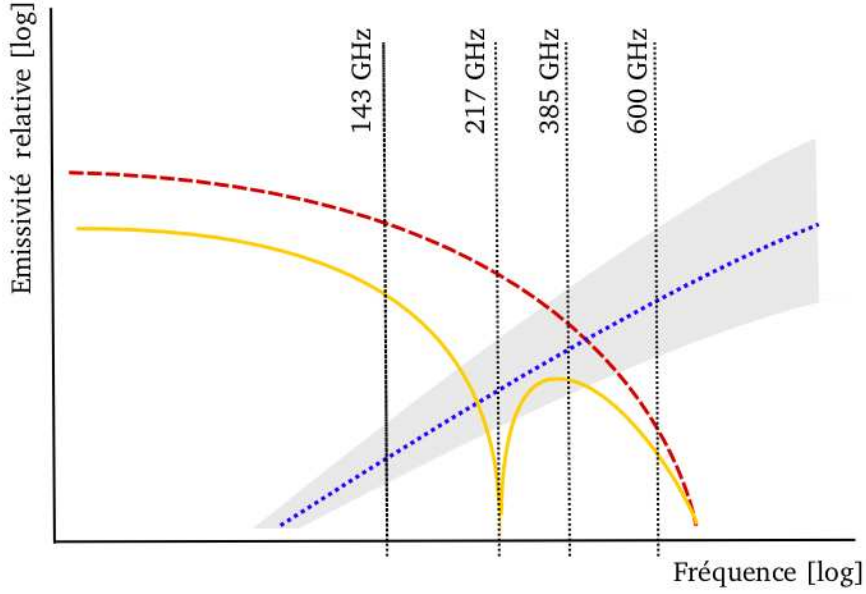


FIG. 4.4.1 – Schéma illustrant les spectres d'émissivité des différentes composantes simulées. La courbe orange en trait plein illustre le spectre de l'effet SZ, la courbe en traits discontinus rouge le CMB, la courbe en pointillés bleus, l'émission thermique de la poussière galactique, tandis que la zone grisée représente le domaine de variation des spectres d'émission des galaxies infrarouges. Les fréquences d'observation des prochaines expériences sont marquées verticalement : 143, 217, 385 et 600 GHz.

Dépendance spectrale de l'effet SZ. On considère dans la simulation des amas non-relativistes, la plupart ont une température inférieure à 10 keV. Dans ce régime l'approximation non-relativiste est justifiée (Rephaeli, 1995), (Rephaeli, 2002a) et la dépendance spectrale du signal SZ, $\Delta I_\nu = I_0 y g(x)$, est alors simplement donnée par (voir équation 2.63) :

$$g(x) = \frac{x^4 e^x}{(e^x - 1)^2} \left[\frac{x(e^x + 1)}{e^x - 1} - 4 \right] \quad (4.11)$$

avec $x = \frac{h\nu}{kT_{CMB}}$ la fréquence réduite d'observation et $I_0 = 2 \frac{(kT_{CMB})^3}{(hc)^2} = 2.28 \times 10^4 \text{ mJy/arcmin}^2$. Cette dépendance spectrale est, dans le cadre de l'approximation

non-relativiste, la même pour tous les amas.

Dépendance spectrale du CMB. Le spectre du monopôle du CMB est celui d'un corps noir de température $T_{CMB} = 2.725$ K :

$$I_\nu = I_0 p(x) \quad (4.12)$$

avec,

$$p(x) = \frac{x^3}{e^x - 1} \quad (4.13)$$

et x la fréquence réduite.

Dépendance spectrale des composantes infrarouges. On suppose que le spectre d'émission des sources infrarouges, galaxies et poussière galactique, est celui d'un corps gris paramétrisé comme suit :

$$F(\nu) = \nu^\alpha \text{Planck}(\nu, T_0). \quad (4.14)$$

où Planck représente le spectre de corps noir de Planck, ν , la fréquence, T_0 la température du corps noir équivalent, et α l'indice spectral.

- Pour les galaxies infrarouges la température de corps noir équivalent est fixé à $T_0 = 30$ K, tandis que l'indice spectral varie entre 1.5 et 2. En réalité les observations montrent que la température varie aussi, entre 20 et 40 K, et de manière corrélée avec l'indice spectral. Nous avons simplifié en fixant T_0 à une valeur moyenne et en tirant aléatoirement α entre 1.5 et 2.
- Pour la poussière galactique la température est fixé à $T_0 = 20$ K et l'indice spectral à $\alpha = 2$. Ces valeurs reproduisent bien les observations à haute latitude galactique (Lagache et al., 1998).

L'émission du CMB et de l'effet SZ est maximum entre 100 et 300 GHz. Pour les sources chaudes que sont les galaxies infrarouges et la poussière galactique (T_0 entre 20 et 40 K), dans la plage de fréquences 100 – 600 GHz, on peut utiliser l'approximation de Rayleigh-Jeans. Ainsi les spectres de corps gris sont dominés par l'indice spectral, la normalisation étant donnée par ailleurs grâce à la carte de Schlegel pour la poussière galactique et aux mesures de comptage des sources extra-galactiques.

Mélange des sources Une carte observée $S_{obs;\nu}$ dans une bande optique centrée sur la fréquence ν peut-être modélisée comme la somme des émissivités des différentes composantes physiques à cette fréquence, en considérant le cas où les propriétés des sources sont homogènes sur les cartes on peut écrire :

$$S_{obs;\nu}(\vec{x}) = \sum_i a_{\nu,i} S_{phy;i} + N_\nu(\vec{x}) \quad (4.15)$$

où \vec{x} indique la direction d'observation, le terme N_ν est la contribution du bruit instrumental à l'observation. $a_{\nu,i}$ est l'émissivité de la composante i à la fréquence ν

et $S_{phy;i}$ décrit les différentes sources physiques émissives dans la plage de fréquence considérée :

$$S_{phy;\nu}(\vec{x}) = \{CMB_\nu(\vec{x}); Gal_\nu(\vec{x}); SZ_\nu(\vec{x}) \dots\} \quad (4.16)$$

En réalité la composante des sources ponctuelles extra-galactiques n'est pas homogène sur la carte. Chacune des galaxies possède sa propre émissivité et est ajoutée individuellement à chacune des cartes aux différentes fréquences :

$$S_{obs;\nu}(\vec{x}) = \sum_{i=SZ, CMB, Poussière} a_{\nu,i} S_{phy;i}(\vec{x}) + \sum_j a_{\nu,IR}(j) S_{phys;IR}(\vec{x}) + N_\nu(\vec{x}) \quad (4.17)$$

où j décrit les sources ponctuelles, i les composantes homogènes, ν la fréquence d'observation et \vec{x} la position considérée. Le terme décrivant les sources infrarouges fait intervenir l'émissivité particulière de chacune des galaxies : $a_{\nu,IR}(j)$.

4.5 Observations multi-fréquences bolométriques

Dans la section précédente nous avons vu comment prédire les émissivités des différentes composantes astrophysiques dans la bande 100 – 600 GHz. Dans cette section je présente le modèle que nous avons mis en place dans la simulation. Bien que nous ayons été guidés par l’expérience Olimpo pour définir ce modèle il reste assez général et est applicable à d’autres expériences comme APEX ou Planck.

Les prochaines expériences multi-fréquences bolométriques observeront le ciel sur des surfaces allant d’une centaine de degrés carrés à tout le ciel pour Planck. A la différence des données des instruments équipés de caméra CCD, les données brutes des expériences bolométriques sont du type TOD, *Time Ordered Data*. La mesure se fait à une fréquence d’échantillonnage donnée. Lors du mouvement du pointage les mesures sont enregistrées à la suite les unes des autres. La reconstruction des cartes à partir des TOD est un travail délicat et constitue le sujet de nombreuses publications (Yvon and Mayet, 2005). Les déconvolutions et premiers filtrages se font à cette étape. Dans notre simulation nous travaillons directement sur les cartes en supposant l’étape du *Map-Making* déjà faite. Je justifie cette approche dans le paragraphe concernant la stratégie de pointage.

La figure 4.5.1 donne un exemple de cartes du ciel générées par la simulation aux fréquences : 143, 217, 385 et 600 GHz, incluant tous les effets instrumentaux modélisés (lobe d’antenne, bruit blanc, bande d’intégration spectrale). On constate que l’effet SZ est sous dominant aux différentes fréquences d’observation : le travail d’extraction des amas est précisé au chapitre 5.

4.5.1 Les fréquences d’observation

Nous nous sommes concentrés dans notre simulation aux quatre fréquences : 143, 217, 385 et 600 GHz. Ce sont des fréquences communes à Olimpo et Planck et les deux premières sont celles d’APEX. Etant optimisées pour la détection de l’effet SZ elles constituent donc un choix raisonnable pour simuler les prochaines expériences de sondage grand champ d’amas SZ.

Le second point important concernant les fréquences d’observation est la largeur des bandes d’intégration pour chacune des fréquences. La sélection en fréquence s’obtient grâce à l’association de cornets et de filtres optiques situés devant les bolomètres. Cet ensemble constitue un filtre passe-bande. Dans la simulation nous prenons en compte ce phénomène en introduisant des fenêtres d’intégration de type “top-hat” de largeur $\Delta\nu$ proportionnelle à la fréquence centrale ν ou constante. Les valeurs utilisées pour les différentes expériences sont données au chapitre 3.

4.5.2 Bruit de mesure

On suppose que l’ensemble des : bruit de photon, sensibilité des bolomètres, bruit de l’électronique de mesure et efficacité des miroirs peut être résumé par un niveau de bruit équivalent en température sur le ciel : B_{eqT} en unité $\mu K_{CMB} s^{1/2}$.

Le modèle de bruit le plus simple est un bruit blanc dans les TOD. On ne tient pas compte dans le modèle de bruit des éventuelles dérives lentes du système : variation

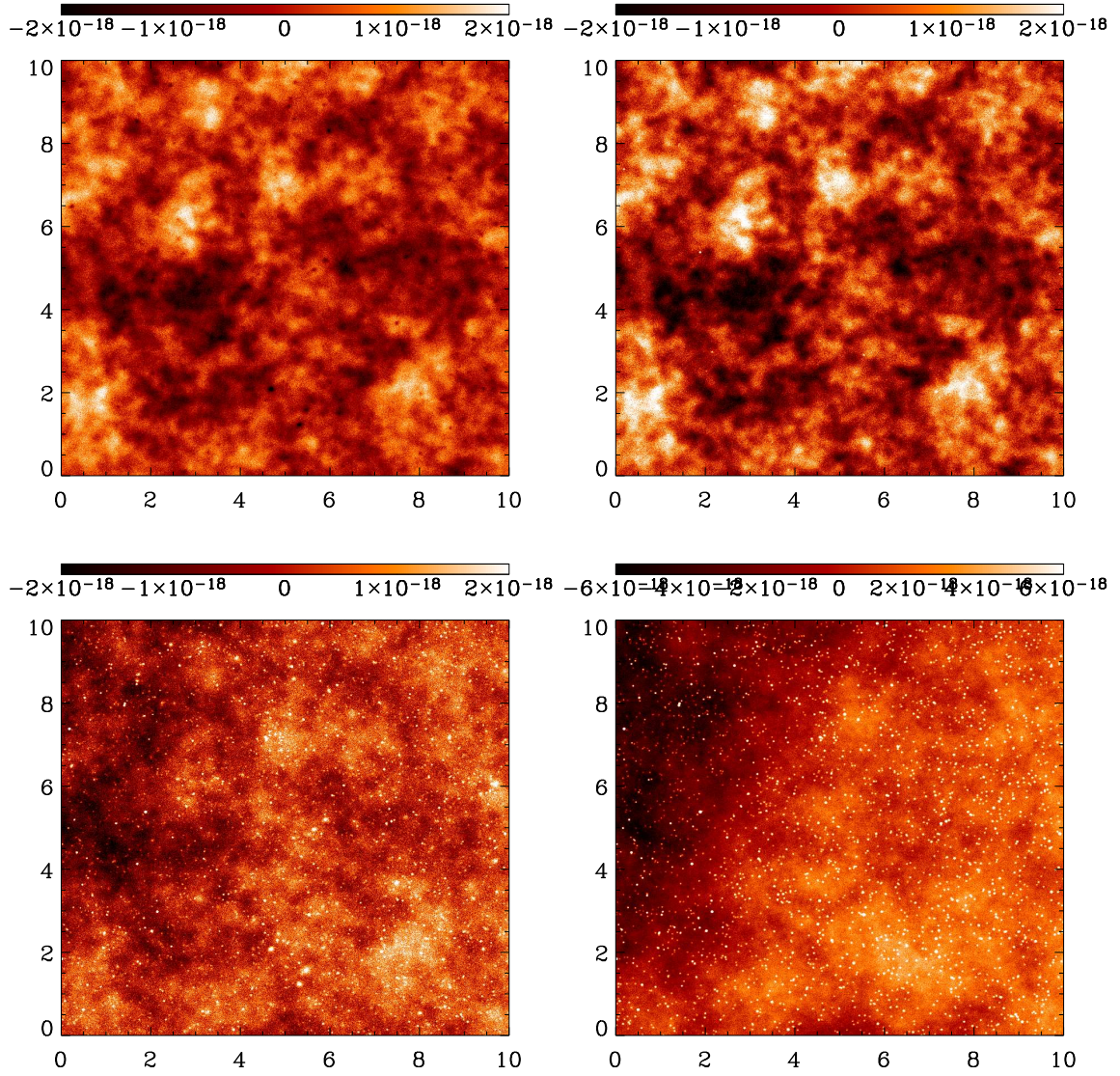


FIG. 4.5.1 – Les cartes observées aux quatre fréquences d'observation communes à Olimpo et Planck, APEX se limitant aux deux fréquences les plus basses. Les effets instrumentaux de bruit et lobe d'antenne sont introduits selon le modèle de l'expérience Olimpo (voir le tableau 3.1). De haut en bas et de gauche à droite : 143, 217, 385 et 600 GHz. Les abscisses et ordonnées sont données en degrés.

de température, changement de sensibilité des bolomètres, etc. Ces phénomènes introduisent un bruit basse fréquences en $1/f$. Le bruit instrumental est ainsi entièrement déterminé par son écart-type σ_N pour chaque pixel. Après projection sur le ciel, le bruit dans chaque pixel est calculé par un tirage aléatoire selon une gaussienne de largeur σ_N :

$$\sigma_N = \frac{B_{eq}T}{\sqrt{N_{bolo}T_{pix}}} \quad (4.18)$$

où N_{bolo} est le nombre de bolomètres à la fréquence considérée et T_{pix} le temps passé à observer chaque pixel. Les niveaux de bruit blanc sont donnés au chapitre 3.

L'utilisation d'un modèle de bruit blanc est justifié par le fait qu'on suppose avoir suffisamment de redondance sur les cartes pour éliminer correctement toute contribution basse fréquence du bruit. Par ailleurs, un filtrage passe-haut des bruits basse fréquence n'aurait que peu d'influence sur le signal SZ. En effet les amas ont des tailles typiques inférieures à 0.2 degrés sur le ciel tandis que le bruit basse fréquence se situe à des échelles plus grandes. Un simple filtrage permettrait de s'en affranchir suffisamment pour l'extraction du signal SZ. Un tel filtrage serait, bien sûr, plus gênant pour l'extraction du signal du CMB ou de la poussière galactique.

4.5.3 Stratégie d'observation

Pour les observations grand champ, Olimpo utilisera une stratégie d'observation dite "à élévation constante". Les 300 deg² seront couverts par juxtaposition de petites portions du ciel. Chacune étant observée avec une élévation par rapport à l'horizon constante. Le miroir primaire sera actionné pour un balayage horizontal dans le référentiel de la nacelle, tandis que le ciel défilera devant le télescope. Un repointage en élévation permettra d'observer la même portion du ciel plusieurs fois et d'augmenter la redondance des mesures. Le fait de travailler à élévation constante permet de simplifier le traitement des effets atmosphériques. En effet ce type d'observations se fait, sur la même portion de ciel, à épaisseur d'air constante. Ainsi on suppose que les effets atmosphériques pourront être corrigés avant l'étape du Map-Making. Dans notre simulation on travaille directement sur les cartes et non sur les TOD. Le fait de travailler directement sur les cartes, plutôt que sur les TOD suppose :

1. que le bruit basse fréquence pourra être supprimé à l'étape du Map-Making,
2. que les effets de l'atmosphère, fortement dépendant de la stratégie de pointage, pourront être éliminés : par un bon choix de stratégie de pointage et grâce aux fréquences d'observation les plus hautes.

4.5.4 Lobe d'antenne et précision de pointage

Les résolutions angulaires s'améliorant, il devient primordial de contrôler la forme du lobe principal de l'instrument et la direction du pointage. La mesure de la forme du lobe se fait souvent au sol et en vol, par pointage de sources ponctuelles bien connues : Jupiter, Saturne ou Mars. Le pointage est contrôlé, dans Olimpo, par quatre points GPS associés à une caméra CCD observant les étoiles dans le champ.

Le fait de travailler sur des cartes et non des TOD nous oblige à adopter un lobe symétrique comme modèle de lobe d'antenne. En pratique nous utilisons une gaussienne, caractérisée par sa largeur à mi-hauteur. Celle-ci est donnée, pour les différentes fréquences d'Olimpo dans le tableau 3.1. En pratique, les cartes observées ont été convoluées par un noyau gaussien dont la largeur à mi-hauteur dépend de la fréquence considérée.

La taille du lobe d'antenne va nous guider pour choisir la taille des pixels des cartes de notre simulation. Afin d'éviter des effets algorithmiques indésirables de numérisation il convient de prendre une taille de pixel au moins trois fois plus faible que la taille du lobe d'antenne. En pratique nos cartes se trouvent largement sur-échantillonnées. Ceci constitue un avantage et un inconvénient : les algorithmes d'extraction que nous avons mis en place (voir le chapitre 5) fonctionnent mieux, mais le niveau de bruit par pixel est plus élevé.

4.6 Conclusion

J'ai décrit dans ce chapitre les différents éléments astrophysiques et instrumentaux inclus dans notre simulation :

- émission du CMB,
- émission des amas par effet SZ,
- émission infrarouge de la poussière galactique et des sources extra-galactiques,
- nombre et largeur des bandes d'intégration des fréquences d'observation,
- lobe d'antenne,
- bruit de mesure.

Les évolutions qu'il me semble le plus souhaitable d'intégrer dans un futur travail sont les suivantes :

1. la corrélation à deux (ou plus) de points dans le placement des amas de galaxies. Je reviendrai sur ce point au chapitre 6 car le fait de ne pas inclure cet effet a certainement une influence sur les algorithmes de détection d'amas que nous avons mis en place. Plusieurs méthodes pourraient être utilisées pour corréler la position des amas entre eux :
 - utiliser des simulations N-corps simplifiées reproduisant la corrélation complète (à 2 points et plus) des surdensités de matière sombre, puis localiser les positions corrélées et ajouter la composante baryonique de la même manière que dans notre simulation, c'est à dire en supposant un profil de densité de King. Cette méthode aurait l'avantage d'être rigoureuse et moins lourde que l'utilisation d'une simulation complète.
 - La méthode "lourde" consisterait à utiliser une simulation N-corps hydrodynamique incluant les baryons. Il serait ensuite possible de projeter les volumes d'univers pour obtenir des cartes 2D d'amas en SZ, X, etc.
 - Une méthode analytique a été développée au PCC par J.B. Melin, J.G. Bartlett et J. Delabrouille (Melin et al., 2005). Elle est basée sur le découpage de l'espace observé dans un angle solide en coquilles de redshift et sur le calcul de la distribution de densité à partir du spectre de densité en fonction du redshift. Les amas sont tirés aléatoirement selon ce spectre de densité au redshift considéré. Cette méthode donne une fonction de corrélation à deux points en accord, dans le régime linéaire, avec les simulations N-corps du Hubble Volume mais ne peut reproduire les fortes non-linéarités associées à l'effondrement des structures.

Les méthodes analytiques devraient permettre d'obtenir la bonne fonction de corrélation à deux points, cependant elles ne reproduisent pas visuellement l'aspect des simulations à N-corps issu des corrélations d'ordre plus élevé : filaments, surdensités fortement non linéaires, etc.
2. Par ailleurs il me semble important, pour simuler le mieux possible les données à attendre des instruments de commencer la simulation directement au niveau des TOD. Ceci nous permettrait d'inclure :
 - les effets atmosphériques de changement de masse d'air lors des repointages,
 - les effets du bruit en $1/f$,
 - les effets d'un lobe d'antenne asymétrique.
3. Enfin il apparaît essentiel, à la vue des résultats concernant les contaminations

par les sources infrarouges (Chapitre 6), d'inclure dans la simulation les sources ponctuelles radios. Ces sources, tout comme les sources infrarouges, possèdent chacune un spectre propre et leur contribution aux observations à 143 GHz et même 217 GHz doit être prise en compte. C'est potentiellement une forte source de contamination des catalogues d'amas SZ reconstruits.

Les orientations choisies pour cette simulation sont clairement tournées vers une production rapide de cartes du ciel observé simulées. Ceci est nécessaire pour pouvoir produire une fonction de sélection simulée d'un instrument multi-fréquences à caméra bolométrique conçue pour effectuer des sondages grand champ sub-millimétriques.

On constate sur la figure 4.5.1 que l'effet SZ est sous dominant aux différentes fréquences d'observation. Le travail que je présente dans le chapitre 5 a pour objet de mettre en place une chaîne d'extraction du signal SZ à partir des cartes observées.

Chapitre 5

Détection d'amas sur les cartes simulées

Sommaire

5.1	Introduction	100
5.2	Isoler l'effet SZ	103
5.2.1	Séparation aveugle de sources	103
5.2.2	JADE	104
5.2.3	Représentation dans une base multi-échelles	105
5.2.4	Résultats	106
5.2.4.1	Séparation en aveugle : JADE avec des données bruitées	106
5.2.4.2	Comment améliorer les choses ?	109
5.2.4.3	Conclusion	109
5.3	Filtrage des cartes	113
5.3.1	Filtrage Gaussien	113
5.3.2	Filtrage non-linéaire : ME-FDR	113
5.3.2.1	Maximum Entropy Method	114
5.3.2.2	Multiscale Entropy	115
5.3.2.3	Filtrage ME-FDR	117
5.3.3	Résultats	118
5.4	Détection des amas	123
5.4.1	Algorithme de détection de sources brillantes	123
5.4.2	Association et marquage des sources	125
5.4.3	Qualité du catalogue reconstruit	127
5.4.4	Résultats	127
5.5	Conclusion	130

5.1 Introduction

LA figure 4.5.1, résultat de la simulation présentée au chapitre 4, illustre toute la difficulté pour extraire le signal SZ qui est sous dominant aux quatre fréquences considérées. Cependant on peut distinguer (voir figure 5.1.1) les amas SZ les plus brillants de manière localisée. On constate cependant qu'un simple seuillage des cartes ne permet pas d'extraire le signal SZ.

L'extraction du signal SZ à partir des cartes observées par les expériences sub-millimétriques est le sujet de publications nombreuses depuis le début des années 2000.

En 2002, à partir de simulations réalistes (dix fréquences, sources galactiques, extra-galactiques, CMB, SZ) de type Planck, D. Herranz publie un article portant sur deux méthodes d'extraction du signal SZ grâce à des filtres multi-fréquences adaptés : SAMF pour *Scale-Adaptative MultiFilter*, et MMF pour *Matched Multi-Filter* (Herranz et al., 2002a,b). Sa méthode est rapide et permet de reconstruire correctement le flux des amas. Cependant elle suppose leur profil de brillance connu. À partir des mêmes simulations, J.M. Diego publie la même année (Diego et al., 2002) une autre méthode qui se décompose en deux étapes : tout d'abord un nettoyage des cartes, puis un filtrage bayésien pour extraire le signal SZ. Le nettoyage des cartes consiste à éliminer, par des méthodes adaptées (seuillage, filtrage, etc.), les différentes contaminations : sources ponctuelles, poussières et CMB. Ensuite le filtrage permet d'augmenter le rapport signal sur bruit de l'effet SZ, le niveau de bruit ayant augmenté à la première étape (compromis nécessaire pour éliminer les contaminations). Le filtrage bayésien utilisé repose sur la connaissance *a priori* du spectre de puissance de l'effet SZ et sur la connaissance de la forme des amas. Ces deux méthodes fonctionnent bien et permettent de retrouver plus de 10000 amas pour les simulations plein ciel de Planck. En 2003 Hobson and McLachlan (2003) publient une méthode efficace de détection d'amas et de reconstruction de leur taille et amplitude basée sur une approche bayésienne et sur un algorithme de type *Monte-Carlo Markov Chain*. Cette méthode s'avère efficace mais extrêmement lente. En 2004, J.-B. Melin (Melin, 2004), reprend l'approche utilisée par D. Herranz des MMF et l'améliore pour éviter les problèmes de confusion de sources grâce à un algorithme de séparation de sources superposées (*deblending*). L'application de cette méthode aux données de WMAP, VSA et SuZIE permet de reconstruire correctement les flux de COMA et A1795 du catalogue d'Abell, considérés comme des sources connues et de position connue également. Le point fort de ce filtrage est sa rapidité de mise en oeuvre et une bonne photométrie de reconstruction du flux total des sources détectées malgré les fortes incertitudes sur leur taille qui dominent les erreurs. En 2005 Pierpaoli et al. (2005) ont publié un algorithme de détection d'amas dans les cartes multi-fréquences de type Planck et ACT (*Atacama Cosmology Telescope*). L'algorithme utilisé est basé sur une méthode bayésienne de filtrage où le signal SZ est décrit par la distribution des coefficients "voisins" de sa décomposition sur une base d'ondelettes. La méthode, moyennant de nombreux *a priori* physiques, se révèle efficace pour la reconstruction de l'amplitude du signal SZ.

Je considère dans mon travail de thèse des cartes du ciel observé par une expérience multi-fréquences bolométrique (typiquement Olimpo), guidé par l'idée que les cartes

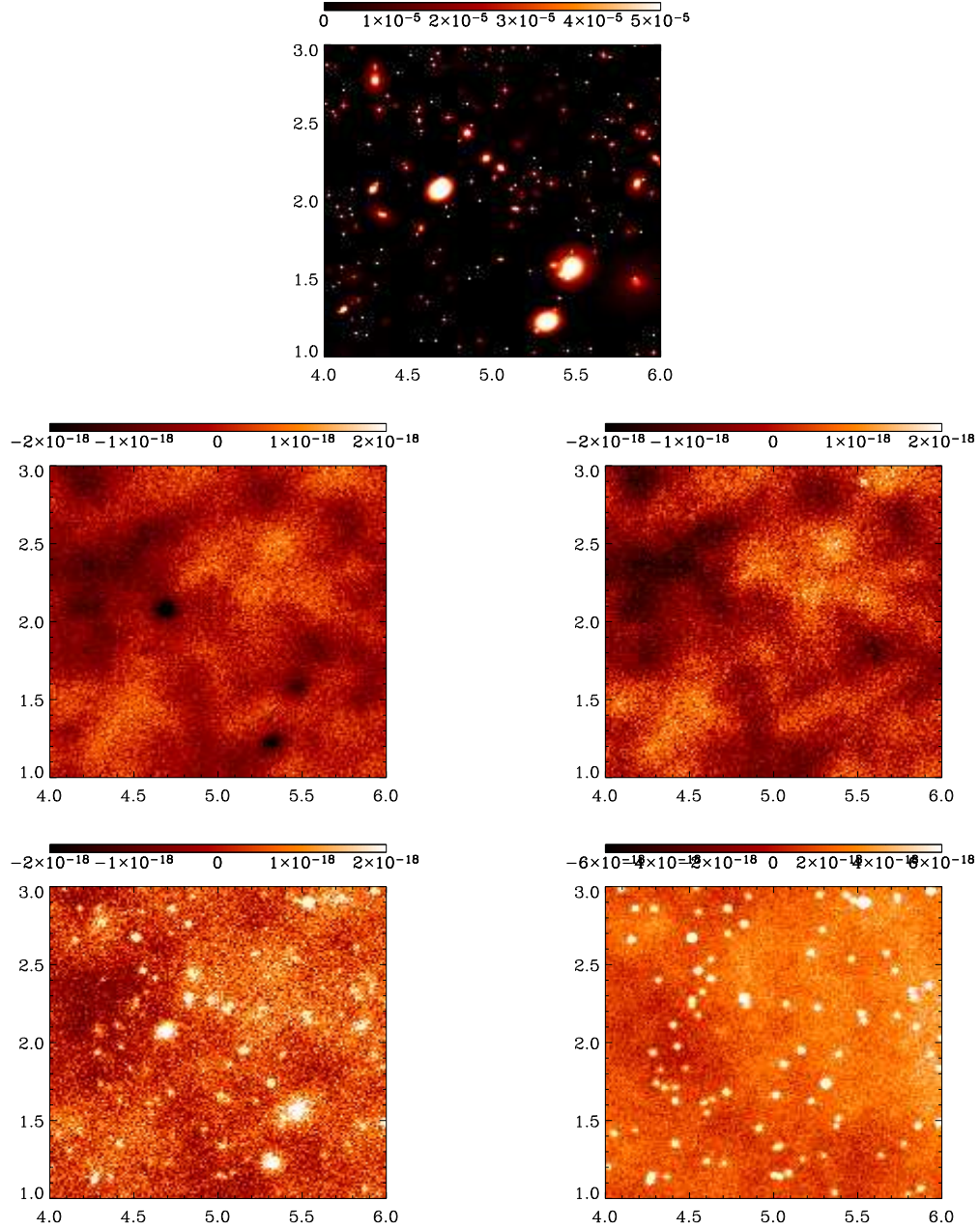


FIG. 5.1.1 – Exemple d’amas de galaxies simulés (figure du haut) et vus aux quatre fréquences (de gauche à droite et de haut en bas) : 143, 217, 385 et 600 GHz dans les conditions expérimentales d’Olimpo données dans le tableau 3.1. Dans la quatrième figure, carte à 600 GHz, les sources brillantes sont des sources infrarouges. A 385 GHz on distingue des amas et des sources infrarouges, à 217 GHz on distingue encore en haut à droite de la carte une galaxie infrarouge. A 143 GHz on voit les amas comme des points froids.

observées aux différentes fréquences sont, essentiellement, des combinaisons linéaires des différentes émissions prises en compte dans la simulation. Nous avons orienté notre travail vers la mise en oeuvre d'une méthode de séparation en composantes indépendantes (ICA, *Independent Component Analysis*) grâce au logiciel JADE, dont je décris le principe dans la section 5.2.1. Aux différentes composantes physiques constituant le signal aux différentes fréquences de l'instrument nous avons ajouté un bruit blanc représentant les bruits d'origine expérimentale (voir 4.5.2). Pour extraire le signal SZ du bruit (que la séparation en composantes indépendantes ne cherche pas à éliminer) nous avons mis en oeuvre une méthode de filtrage non-linéaire fondée sur une méthode de maximisation d'entropie (5.3.2), MEM (Maximum Entropy Method) que je décrirai par la suite. Ces différentes étapes ont permis d'obtenir une carte de signal SZ débruitée. J'expliquerai ensuite la méthode que nous avons mise en place pour extraire un catalogue d'amas à partir de cette carte SZ reconstruite (voir la section 5.4) afin de calculer une fonction de sélection de la chaîne instrumentale d'observation. En comparant ce catalogue estimé au catalogue simulé ayant servi à la construction de la carte du ciel, je serai en mesure d'évaluer la qualité de la chaîne de détection et de la comparer à d'autres méthodes plus standard en termes de filtrage. Ce travail fait l'objet d'un article soumis à *A&A* écrit en collaboration avec D. Yvon, S. Pires, Y. Moudden, S. Anthoine et E. Pierpaoli. Ce chapitre répond aux questions :

1. Comment isoler l'effet SZ ?
2. Comment s'affranchir du bruit instrumental ?
3. Comment extraire de la carte reconstruite les amas en tant qu'objets ?

Dans la première section je m'attache à montrer comment, à partir des cartes observées, il est possible d'extraire les composantes physiques initiales et d'isoler l'effet SZ. Je montre que JADE est bien adapté à l'extraction du signal SZ, notamment quand les signaux observés sont décomposés sur une base d'ondelettes multi-échelles. Je montrerai les résultats préliminaires obtenus avec cet algorithme. Dans la section suivante j'introduirai la notion de filtrage des cartes en insistant sur les filtres non-linéaires et l'approche bayésienne pour en arriver à la méthode ME-FDR que nous avons mise en oeuvre. Enfin dans la dernière section je présenterai la méthode employée pour passer d'un signal SZ (cartes reconstruites de paramètre de Compton) à des objets SZ (catalogue d'amas). Ceci me permettra de quantifier de manière complète la qualité des différents algorithmes mis en oeuvre dans cette chaîne de détection de sources SZ. Un schéma résumant les différentes étapes de la méthode est donné en conclusion du chapitre (5.5.1).

5.2 Isoler l'effet SZ

L'un des critères qui a guidé la mise en place de la chaîne d'extraction d'amas SZ à partir des données est de parvenir à un algorithme qui puisse travailler "en aveugle". Nous n'introduisons dans la chaîne de traitement aucun *a priori*, ni sur le signal SZ que nous cherchons à retrouver, ni sur les contaminants. Dans cette optique nous avons cherché des méthodes de séparation aveugle de sources, BSS pour *Blind Source Separation*, dont je présente le principe dans le paragraphe suivant.

5.2.1 Séparation aveugle de sources

Le traitement de signaux multi-fréquences dans les applications astrophysiques cherche à exploiter les corrélations spatiales des cartes relatives à différentes fréquences ν pour extraire les signaux "physiques" : spectre de puissance du CMB, amas de galaxies, etc. Bien que l'approximation d'une émissivité homogène des sources sur l'ensemble de la carte ne soit pas tout à fait vérifiée dans le cas des galaxies infrarouges (4.4) elle l'est pour les autres composantes au niveau de sensibilité des expériences étudiées ici. En supposant que les lobes optiques instrumentaux sont indépendants de la fréquence d'observation ν , on peut supposer observer un mélange linéaire de sources physiques :

$$S_{obs}(\vec{x}) = A S_{phys}(\vec{x}) + N(\vec{x}) \quad (5.1)$$

où S_{obs} et S_{phys} sont des vecteurs de taille : $m \times 1$, $n \times 1$ et A est une matrice de taille $m \times n$. La multiplication de S_{phys} par A représente le mélange linéaire des n sources statistiquement indépendantes pour former m signaux mesurés. Les observations sur les quatre fréquences ne sont pas incohérentes : cette modélisation linéaire permet de tenir compte, lors du traitement des signaux, de ces cohérences. En supposant que A soit complètement inconnue, retrouver les différentes composantes physiques à partir des canaux d'observations peut-être vu comme un problème de séparation aveugle de sources (*Blind Sources Separation*, BSS). Le but des méthodes BSS est de retrouver, à partir de signaux observés S_{obs} (ici des cartes du ciel à m différentes fréquences), n signaux physiques ou sources S_{phys} .

Les méthodes ICA ont été développées pour résoudre ce type de problème de séparation aveugle de sources, c'est à dire estimer, à partir d'une observation S_{obs} , la matrice de mélange A afin de retrouver les signaux S_{phys} . Il est important de remarquer que les méthodes ICA permettent de retrouver les sources indépendantes à un facteur d'échelle et une permutation près. En pratique nous serons confrontés à ce fait. J'explique dans le paragraphe sur les résultats de la méthode ICA comment traiter cette difficulté.

Les méthodes ICA reposent principalement sur l'indépendance statistique des sources S_{phys} . Cette hypothèse d'indépendance statistique peut paraître forte mais elle est en pratique presque toujours vraie. Les méthodes ICA vont plus loin que les méthodes d'analyse en composantes principales (PCA, *Principal Component Analysis*) : en effet une simple décorrélation ne suffit généralement pas à retrouver les sources S_{phys} . Le critère d'orthogonalité (recherché par les méthodes PCA) des sources n'est pas suffisant, dans notre cas, pour les séparer.

Les algorithmes de séparation de composantes et d'estimation des matrices de mélange dépendent du modèle utilisé *a priori* pour décrire la distribution de probabilité des sources. Certaines méthodes se placent dans le cas de signaux non-bruités et dont les distributions de probabilités sont non-gaussiennes (soit toutes, soit une pouvant être gaussienne). C'est le cas de la plupart des méthodes disponibles : JADE (Cardoso, 1999), FastICA, Infomax (Hyvarinen et al., 2001). D'autres méthodes cherchent à séparer des composantes dont les distributions de probabilités sont supposées gaussiennes. La séparation se base alors sur le fait que dans une représentation donnée (Fourier, espace réel, ondelettes, etc.) les sources ont des profils de variances non-proportionnels. C'est le cas de la méthode SMICA, *Spectral Matching ICA method* (Delabrouille et al., 2003; Moudden et al., 2005) travaillant dans l'espace de Fourier. Des méthodes ICA plus standards ont été utilisées avec succès pour extraire le spectre du CMB dans l'analyse des données des expériences multi-fréquences futures ou actuelles telles que Archeops, WMAP ou Planck (Delabrouille et al., 2003; Maino et al., 2002; Kuruoglu et al., 2003).

Dans notre cas le signal SZ est fortement non-gaussien et les méthodes de la première catégorie conduisent à de meilleurs résultats. Nous avons utilisé JADE pour effectuer la séparation en composantes indépendantes des cartes observées. Le paragraphe suivant explique les grands principes de son fonctionnement.

5.2.2 JADE

La méthode mise en oeuvre dans JADE, *Joint Approximate Diagonalization of Eigenmatrices method*, suppose un mélange linéaire des sources S_{phys} (voir l'équation 5.1) dont les distributions de probabilités sont issues de processus aléatoires non-gaussien *i.i.d.*. Le terme *i.i.d.*, *independently and identically distributed*, signifie deux choses :

1. que chaque source S_{obs} à une position donnée est indépendante de S_{obs} aux autres positions,
2. et que, à une position donnée, les différentes sources sont indépendantes les unes des autres.

Ce critère *i.i.d.* n'est pas vérifié dans notre cas : un pixel appartenant à un amas est entouré de "voisins" appartenant aussi à l'amas et leur flux sont corrélés.

Dans notre simulation nous n'avons pas utilisé cette hypothèse pour tous les signaux : les galaxies infrarouges possèdent chacune leur propre dépendance spectrale et les cartes simulées ne sont pas un mélange linéaire des quatre phénomènes physiques (voir l'équation 4.17). Nous verrons que JADE traite efficacement ce problème et parvient tout de même à extraire le signal SZ.

La méthode utilisée dans JADE est conçue pour fonctionner à haut rapport signal sur bruit : $\text{Var}(AS_{phys}) \gg \text{Var}(N)$. La matrice de mélange est supposée carrée et inversible de manière à ce que le mélange corresponde simplement à un changement de base. Bien que la supposition d'un haut rapport signal sur bruit ne soit en pratique pas toujours vérifiée (notamment dans les observations bolométriques) l'algorithme peut être appliqué et un changement approprié de représentation du signal observé peut permettre de minimiser l'impact du bruit comme nous allons le voir par la suite.

Les méthodes PCA travaillent sur des statistiques du second ordre à travers la diagonalisation de la matrice de covariance. Le résultat obtenu est une décorrélation des signaux observés. Cependant les grandeurs statistiques du second ordre ne contiennent pas assez d'information pour séparer les sources. Les méthodes PCA ne permettent généralement pas d'identifier les sources indépendantes correctement. Cependant ce travail de décorrélation constitue la première étape des méthodes dites "orthogonales" dont fait partie JADE. Dans les algorithmes "orthogonaux" on cherche, parmi toutes les bases dans lesquelles les données sont décorréliées (ce qui est le résultat d'un algorithme PCA), celle où les signaux physiques sont les plus indépendants les uns des autres. Ces algorithmes se décomposent donc en deux étapes : après un "blanchiment" du bruit obtenu en multipliant les signaux observés par l'inverse de la matrice de covariance W , on va chercher une rotation R (laissant les signaux décorréliés) qui maximise l'indépendance des sources :

$$\tilde{S}_{phys} = W^{-1} Y = W^{-1} R S_{obs}^{white} = W^{-1} R W S_{obs} \quad (5.2)$$

où \tilde{S}_{phys} est l'estimation des sources et $W^{-1} R W = \tilde{A}^{-1}$ est l'estimation de la matrice de mélange. JADE se base sur ce principe et, comme la plupart des méthodes ICA, exploite les moments d'ordres supérieurs des signaux pour maximiser l'indépendance des sources. La fonction "mesurant" l'indépendance des sources, basée sur un calcul de moment d'ordre quatre, est la suivante :

$$\mathcal{J}_{jade}(R) = \sum_{ij} \sum_{k \neq l} F_{ijkl}^2 \quad (5.3)$$

où i, j, k et l décrivent les différentes composantes y_i de $Y = R W S_{obs}$. Par définition :

$$F_{ijkl} = \mathcal{E}(y_i y_j y_k y_l) - \mathcal{E}(y_i y_j) \mathcal{E}(y_k y_l) - \mathcal{E}(y_i y_l) \mathcal{E}(y_j y_k) - \mathcal{E}(y_i y_k) \mathcal{E}(y_j y_l) \quad (5.4)$$

avec \mathcal{E} la moyenne statistique calculée pour chaque signal d'entrée y_i . Le changement de base (la rotation) donnant la plus grande indépendance entre les signaux est obtenu en minimisant la fonction $\mathcal{J}_{jade}(R)$. Pour plus de détails sur le fonctionnement de JADE consulter Cardoso (1999), Cardoso (1998) et Hyvarinen et al. (2001).

5.2.3 Représentation dans une base multi-échelles

L'utilisation des ondelettes dans l'étape de séparation en composantes se justifie en considérant la remarque faite dans le paragraphe précédent à propos du rapport signal sur bruit. JADE suppose localement que le rapport signal sur bruit est élevé. Un moyen d'augmenter ce rapport consiste à utiliser une représentation appropriée du signal afin de le rendre "plus" i.i.d. La décomposition dans une base d'ondelettes du signal SZ, donnée par ses coefficients, possède une structure particulière dont l'énergie est localisée à la position des amas et est présente à quelques échelles angulaires seulement. L'information contenue dans un coefficient d'ondelette regroupe l'information de position mais aussi de taille sur la carte. Dans notre méthode nous utilisons la décomposition du signal observé S_{obs} dans la base des ondelettes de

Mallat. La transformation inverse après le passage de JADE redonne les sources indépendantes. En effet la décomposition de Mallat est linéaire et n'affecte ni la matrice de mélange A ni la structure du mélange (le modèle linéaire utilisé) :

$$S_{obs}^w = \text{Mallat}(S_{obs}) \quad (5.5)$$

$$\tilde{S}_{phys}^w = (W^{-1} R W) S_{obs}^w \quad (5.6)$$

$$\tilde{S}_{phys} = \text{Mallat}^{-1}(\tilde{S}_{phys}^w) \quad (5.7)$$

où Mallat représente la décomposition multi-résolution de Mallat et l'indice w indique une quantité décomposée dans cette base.

Bien que la décomposition soit linéaire les distributions statistiques des coefficients décrivant les différents signaux physiques ("utiles" ou contaminants) dans la nouvelle représentation sont différentes : les ondelettes permettent de localiser plus précisément les signaux de structures compactes, elles amplifient le signal en utilisant la cohérence entre pixels des structures. Le rapport signal sur bruit dépend de la représentation utilisée, dans le cas des amas il augmente car le bruit reste dispersé dans la représentation en ondelettes tandis que le signal SZ est concentré. La transformation de Mallat est unitaire, l'énergie totale est conservée, i.e. la variance "globale" du signal est inchangée, cependant elle permet de changer la variance "locale" du signal, c'est à dire, qu'elle peut modifier la détectabilité des structures.

Bien que les méthodes ICA orthogonales telles que JADE soient conçues pour traiter des données exemptes de bruit, elles fonctionnent bien avec nos données simulées bruitées. Le passage en représentation par ondelettes rend JADE plus robuste vis-à-vis du bruit. Une fois le calcul de la matrice de mélange effectué selon cette méthode on retrouve le signal des sources par transformation inverse.

5.2.4 Résultats

Je présente dans ce paragraphe les résultats obtenus pour la séparation en aveugle des différentes sources astrophysiques avec JADE en me concentrant sur la restauration du signal SZ. Je répondrai aux deux questions suivantes :

1. Comment se comporte JADE vis à vis de données bruitées ?
2. Quels moyens peut-on utiliser pour améliorer les choses ?

5.2.4.1 Séparation en aveugle : JADE avec des données bruitées

Dans un premier temps nous utilisons JADE, sur les cartes observées aux quatre fréquences, sans pré-traitement des données. La figure 5.2.1 montre les signaux physiques reconstruits : $\tilde{S}_{phys} = \{\tilde{S}_{SZ}; \tilde{S}_{CMB}; \tilde{S}_{Dust}; \tilde{S}_{IR}\}$. A la vue de ces cartes on peut faire les remarques suivantes :

1. on identifie facilement les différentes composantes,
2. le CMB et les amas sont bien reconstruits et séparés,
3. les composantes infra-rouges sont regroupées dans une seule carte,
4. les composantes reconstruites sont bruitées,

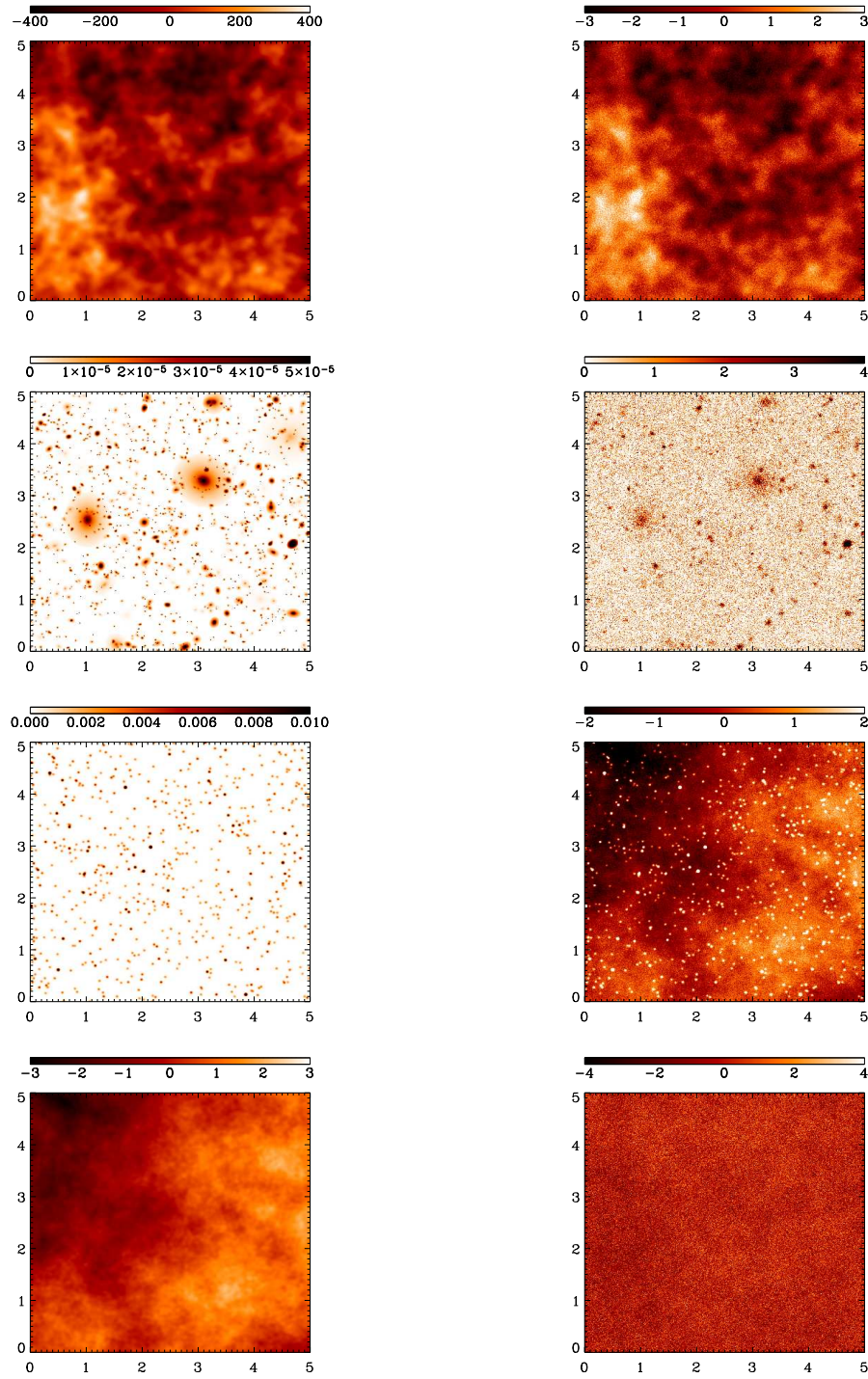


FIG. 5.2.1 – La colonne de gauche montre les composantes physiques initiales : S_{phys} . En haut, le CMB, en-dessous, les amas SZ, ensuite, les galaxies IR, enfin la poussière galactique. La colonne de droite montre les cartes estimées des différentes composantes physiques obtenues par JADE : \tilde{S}_{phys} . En haut, apparaît principalement la composante CMB, en dessous la composante SZ, ensuite un mélange des galaxies IR et de la poussière et enfin, une carte contenant principalement du bruit. Les abscisses et ordonnées sont données en degrés.

5. une des cartes semble ne comporter que du bruit,
6. comme on s'y attendait, les cartes ont perdu leur normalisation.

JADE semble bien adapté à l'extraction du signal SZ et fonctionne sur des données bruitées.

On cherche maintenant à estimer la qualité de la reconstruction de la carte SZ. Nous allons comparer directement la carte reconstruite du signal SZ avec la carte simulée d'effet SZ.

Cette méthode suppose que la carte reconstruite de l'effet SZ soit étalonnée. On utilise pour étalonner la carte reconstruite des amas, un outil de photométrie classique : SExtractor (Bertin and Arnouts, 1996) (utilisé en mode FLUX_AUTO) afin de déterminer rapidement le paramètre de Compton intégral des 100 détections les plus brillantes. On normalise les cartes reconstruites sur ces 100 détections en utilisant leur flux intégral tiré du catalogue simulé des amas. Je reviens plus loin dans ce chapitre sur l'utilisation de SExtractor. En pratique cette méthode obligera à effectuer un pointage sur un ou plusieurs amas brillants connus par ailleurs afin d'étalonner la carte reconstruite. Un étalonnage des bolomètres sur une source ponctuelle comme Jupiter ne suffit pas dans notre cas car la carte à normaliser contient *uniquement* de l'effet SZ : on doit donc étalonner l'algorithme sur une source SZ. Les normalisations fonctionnent très bien : après étalonnage l'erreur sur les paramètres de Compton intégré des 100 sources brillantes est inférieure à 4% pour des sources avec $Y > 2.5 \cdot 10^{-10}$ sr soit $3 \cdot 10^{-3}$ arcmin².

On évalue la qualité des cartes reconstruites en comparant la composante SZ reconstruite à la carte SZ simulée. En pratique on regarde le comportement statistique de la carte d'erreur E :

$$E = \tilde{S}_{SZ} - S_{SZ} \quad (5.8)$$

Une première étape pour évaluer la qualité de la reconstruction consiste à calculer la distorsion D du signal, c'est à dire l'écart-type de la carte d'erreur rapporté à l'écart-type de la carte physique simulée :

$$D = \frac{\sigma(E)}{\sigma(S_{SZ})} \quad (5.9)$$

La distorsion D est une version simplifiée (car nous travaillons avec des signaux reconstruits normalisés aux signaux physiques simulés) de la distorsion relative D_1 introduite par Gribonval et al. (2002). Une information plus riche peut-être obtenue en regardant le comportement statistique de la carte d'erreur en fonction d'une échelle l sur la carte. En pratique on utilise une décomposition multi-échelles en ondelettes de la carte d'erreur et de la carte simulée :

$$E^l = \tilde{S}_{SZ}^l - S_{SZ}^l \quad (5.10)$$

où l indique l'échelle considérée. La distorsion est alors donnée par :

$$D^l = \frac{\sigma(E^l)}{\sigma(S_{SZ}^l)} \quad (5.11)$$

Grâce à ce critère on peut voir quelle partie de la carte est bien reconstruite : plus D (ou D^l) est proche de zéro mieux le signal est reconstruit. Le résultat du calcul de D^l pour JADE appliqué sur les données observées bruitées est donné dans la figure 5.2.3.

5.2.4.2 Comment améliorer les choses ?

JADE est conçu pour traiter des données non bruitées mais fonctionne bien avec nos cartes bruitées pour reconstruire le signal SZ. Cependant nous avons cherché à améliorer les choses. Nous avons suivi pour cela deux pistes :

1. une représentation en ondelettes (décomposition multi-résolution de Mallat) pour les signaux observés : $S_{obs}^w = \text{Mallat}(S_{obs})$,
2. un pré-filtrage des cartes observées pour augmenter le rapport signal sur bruit : $S_{obs}^f = f(S_{obs})$.

Utilisation de la représentation de Mallat. La décomposition en ondelettes de Mallat apporte une amélioration aux grandes échelles du signal. Ceci est illustré par la figure 5.2.2. On remarque sur cette figure que la carte d'erreur établie sans utiliser la décomposition en ondelettes présente beaucoup plus de structures aux échelles considérées. Ces dernières ne semblent pas correspondre à des structures du CMB ou de la poussière galactique (voir les figure 4.1.1). Le résultat du calcul de D^l pour JADE utilisé sur les données décomposées dans la base de Mallat est donné dans la figure 5.2.3.

Pré-filtrage des cartes observées. Afin “d’aider” JADE à estimer \widetilde{A}^{-1} on pré-filtre les données par une fenêtre gaussienne dont le FWHM vaut 2.35 arcmin, soit la moyenne des lobes d’antenne instrumentaux utilisés dans la simulation. On augmente ainsi le rapport signal sur bruit (S/N) dans les données S_{obs} ce qui permettra à JADE d’estimer plus correctement les coefficients de la matrice de mélange \widetilde{A}^{-1} . Une fois cette matrice estimée, on applique la transformation aux signaux bruités originaux afin de retrouver la carte reconstruite d’effet SZ : \widetilde{S}_{SZ} . Les résultats des calculs de D^l pour JADE utilisé sur les données avec un pré-filtrage gaussien (avec ou sans décomposition dans la base de Mallat) sont donnés dans la figure 5.2.3.

5.2.4.3 Conclusion

La figure 5.2.3 compare la qualité des différentes améliorations que nous avons apportées à JADE. L’estimation de la qualité d’une séparation en composantes indépendantes est délicate et a fait l’objet de publications récentes (Gribonval et al., 2002) dont on pourra utiliser les critères pour évaluer la qualité de la reconstruction du CMB, des sources infrarouges extragalactiques et de la poussière.

On peut faire plusieurs remarques à la vue de ces résultats.

1. Les plus petites échelles, 0 et 1 (~ 1 et 2 pixels), sont celles qui récupère tout le bruit instrumental. Il est donc normal que l’on y observe des erreurs plus grandes. Ce phénomène est amplifié par le fait que nous avons sur-échantillonné

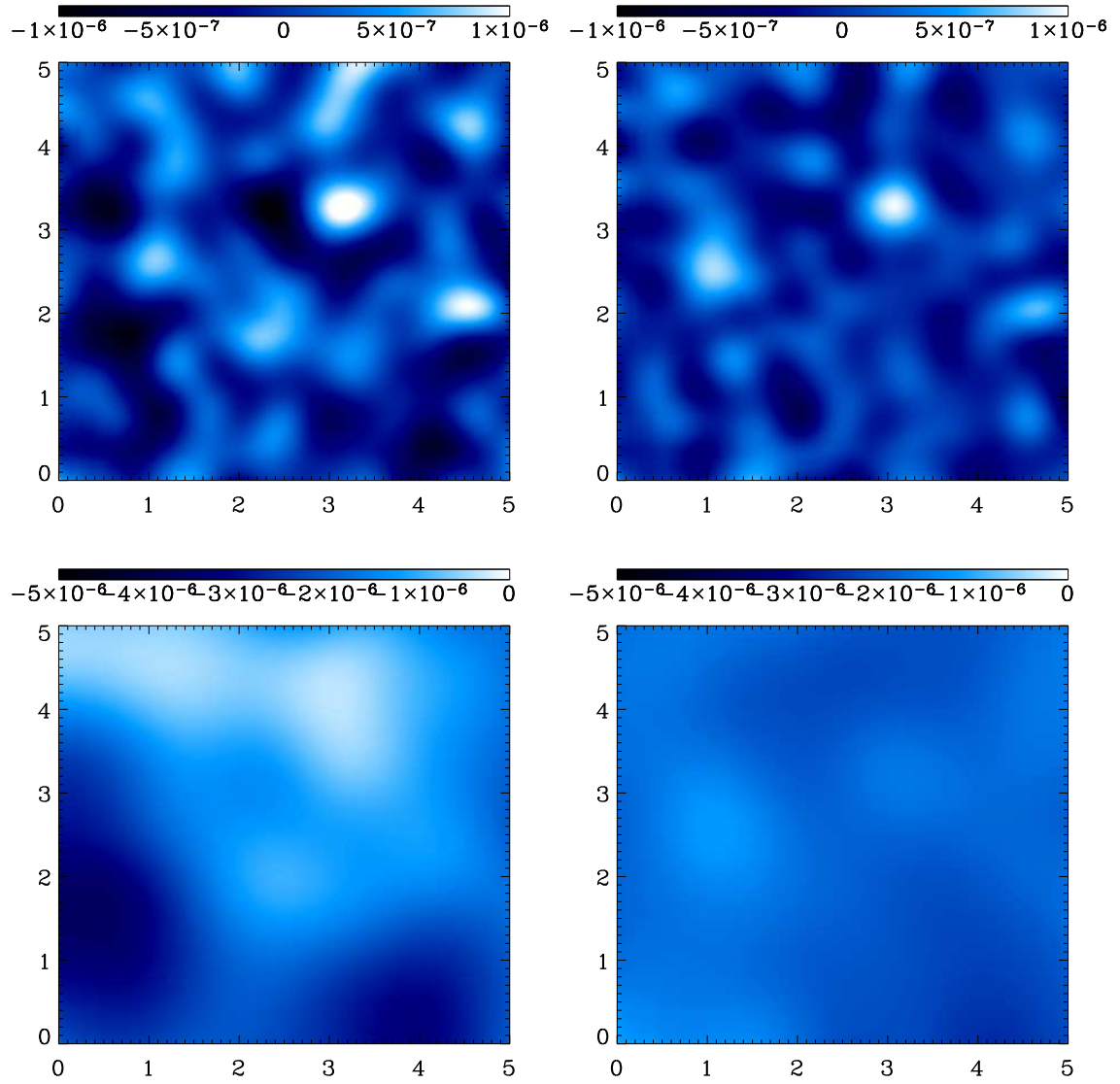


FIG. 5.2.2 – Cartes d'erreurs E^l pour, en haut : $l = 5$ et en bas $l = 6$ et, à gauche sans décomposition dans la base de Mallat, à droite avec décomposition dans la base de Mallat. Ces deux échelles correspondent aux grandes structures de la carte.

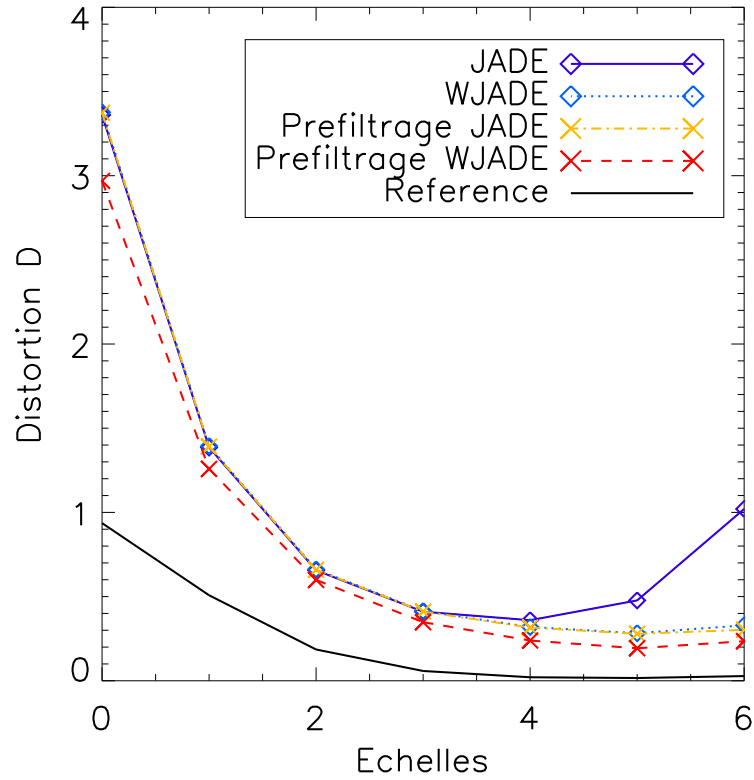


FIG. 5.2.3 – Distorsion D^l (équation 5.11) du signal SZ reconstruit en fonction de l'échelle sur la carte. L'échelle 0 correspond à une taille de 1,2 pixel tandis que l'échelle 6 correspond à une taille d'une soixantaine de pixels. Les cartes de la figure 5.2.2 donnent une idée de la décomposition multi-échelles effectuée. La légende donne la signification des différentes courbes : la courbe "JADE" correspond à l'application directe de JADE sur les données bruitées, "WJADE" indique une décomposition des données dans la base des ondelettes Mallat avant d'appliquer JADE, "Prefiltrage" indique que nous avons pré-filtré les données (comme il est décrit dans le paragraphe précédent) avant d'appliquer JADE. La courbe "Reference" est expliquée dans le paragraphe suivant.

les cartes (voir 4.5.4) : le bruit par pixel est d'autant plus grand que la taille du pixel est plus faible. L'étape suivante de filtrage de la carte SZ reconstruite en apparaît d'autant plus essentielle. Je réutiliserai, dans la section suivante (5.3), l'information D^l pour comparer la qualité des différents filtrages.

2. L'utilisation des ondelettes permet, comme attendu, de "mieux" décrire les différents signaux observés : le SZ mais aussi les structures de grande taille (CMB et poussière). JADE a donc plus de facilité à séparer les différentes composantes et on observe ainsi une meilleure reconstruction du signal SZ aux grandes échelles.
3. Le pré-filtrage permet d'améliorer encore les performances de la séparation, et ce, à toutes les échelles de la carte.

La courbe notée "*Reference*" de la figure 5.2.3 correspond à l'erreur obtenue quand on compare la carte SZ simulée à cette même carte simulée mais convoluée par la fenêtre gaussienne utilisée pour le pré-filtrage et représentative du lobe instrumental. Si la séparation était parfaite et si on arrivait à enlever complètement le bruit des données (sans modifier celles-ci) on obtiendrait cette courbe de distorsion. Le fait qu'elle ne soit pas nulle, particulièrement aux basses échelles, illustre le simple fait que nous ne déconvoluons pas nos cartes du lobe instrumental : le meilleur signal SZ que l'on puisse retrouver sera au mieux équivalent à la carte simulée convoluée. Le but de notre chaîne de détection : ICA puis filtrage est de nous approcher au plus de cette courbe.

5.3 Filtrage des cartes

Nous avons utilisé une méthode de séparation en composantes indépendantes basée sur l'utilisation d'algorithmes BSS pour séparer les différents signaux physiques présents dans les cartes observées. Nous obtenons donc, après cette étape, quatre cartes contenant : le signal de l'effet SZ, le CMB et les composantes IR (galactiques et extragalactiques). Nous nous intéressons maintenant à la carte contenant le signal SZ. Afin d'augmenter le rapport signal sur bruit nous avons testé différents filtrages. Le but du filtrage est d'éliminer au maximum les perturbations tout en respectant l'intégrité du signal original : les éléments utiles de l'image (les amas) doivent être préservés au mieux. Les premiers, et les plus simples, des filtres se fondent sur des algorithmes de convolution linéaire par un "noyau" généralement homogène, mais les limitations de ces techniques (perte des contours des objets, diminution du contraste, etc.) ont conduit au développement de méthodes non-linéaires de filtrage.

Le filtrage permettant l'extraction du maximum d'amas est le filtrage ME-FDR non-linéaire présenté dans la section 5.3.2. Dans un premier temps nous avons testé le filtrage linéaire Gaussien.

5.3.1 Filtrage Gaussien

Le filtrage linéaire le plus simple est le filtrage gaussien homogène. La méthode standard consiste à convoluer la carte SZ reconstruite S_{phys} par une fenêtre gaussienne G ayant une déviation standard σ_G :

$$S_{phys}^G = G * S_{phys} \quad (5.12)$$

La qualité du filtrage dépend fortement de la valeur de σ_G . Cette valeur est fixée *a priori*. Nous convolons la carte par une fenêtre gaussienne de largeur σ_G adaptée au lobe d'antenne de l'instrument. Nous étudions les performances du filtrage pour trois largeurs à mi-hauteur (FWHM) de gaussienne : 2, 3 et 2.35 arcmin correspondant, pour la première au lobe instrumental à 143 GHz, pour la seconde au lobe instrumental à haute fréquence, et la dernière pour la moyenne des largeurs des lobes aux quatre fréquences.

5.3.2 Filtrage non-linéaire : ME-FDR

En plus du filtrage linéaire nous avons testé un filtrage non-linéaire afin d'améliorer encore le rapport signal sur bruit, S/N, des cartes SZ obtenues. Nous ne nous sommes pas attachés, par ce filtrage, à reproduire le profil ou l'amplitude centrale des amas : notre objectif a été l'extraction d'un nombre maximum d'amas. On peut imaginer réutiliser dans un second temps les cartes bruitées pour estimer les caractéristiques des amas connaissant leur position trouvée grâce au filtrage ME-FDR. Pour le moment nous utilisons directement les cartes débruitées obtenues par ce filtrage pour réaliser la photométrie des amas (voir le chapitre 6).

Une des méthodes de filtrage non-linéaire possible est l'utilisation d'un estimateur MAP, *Maximum A-posteriori Probability*. Cette méthode propose de filtrer les données en minimisant une fonction dite "d'énergie" $\mathcal{U}(S_f/S_n)$, où S_n est une

réalisation bruitée du signal que l'on cherche à extraire : dans notre cas c'est la carte \tilde{S}_{SZ} obtenue avec JADE, et S_f est la carte filtrée. Cette énergie peut s'écrire sous la forme :

$$\mathcal{U}(S_f/S_n) = \frac{\|S_n - S_f\|^2}{2\sigma_n^2} + \beta P_r(S_f) \quad (5.13)$$

où σ_n est la déviation standard du bruit dans la carte S_n . Ce bruit est supposé gaussien (ce qui est rigoureux dans notre simulation). Le premier terme est le terme *log-vraisemblance*, il décrit la relation générale entre les signaux S_f et S_n . Le second terme est un terme de *régularisation* où $P_r(S_f)$ représente un *a priori* concernant le signal filtré que l'on souhaite obtenir : S_f . Il existe différentes possibilités pour définir $P_r(S_f)$, les trois plus utilisées sont les suivantes :

- minimisation d'énergie : $P_r(S_f) = \|S_f\|^2$, on cherche alors à retrouver un signal filtré proche du signal bruité mais qui possède une énergie minimum,
- régularisation sur les gradients de l'image pour effectuer un lissage global des données,
- *a priori* entropique : $P_r(S_f) = S_f \log S_f$, on cherche alors à minimiser l'entropie (à définir) du signal filtré. C'est cette approche que l'on utilise dans la suite ; il existe énormément d'approches possibles.

Le filtrage ME-FDR, *Multiscale Entropy using False Discovery Rate*, est basé sur une méthode de maximum d'entropie (MEM, *Maximum Entropy Method*), étendue au cas multi-échelles (ME), d'un signal représenté dans une base d'ondelettes suivi par une méthode automatique de seuillage (FDR). Le paramètre de régularisation qui en résulte (voir l'équation 5.29) est de type entropique pour un signal multi-échelles, limité à un support "utile". Le filtrage qui en résulte s'adapte de manière locale, spatialement et vis-à-vis de l'échelle, au bruit présent. Je décris ici ces différentes étapes.

5.3.2.1 Maximum Entropy Method

Les méthodes de maximum d'entropie sont utilisées fréquemment en astronomie pour le traitement des images (voir Starck et al. (2001); Marshall et al. (2002); Starck and Murtagh (2002) pour une description complète). Elles sont basées sur des méthodes bayésiennes de calcul d'entropie. L'approche Bayésienne fournit un moyen d'incorporer des connaissances *a priori* dans l'analyse de données. Le choix de l'*a priori* est un point délicat de ces méthodes. Dans notre méthode un *a priori* entropique est utilisé. L'entropie est un indicateur du contenu en information d'un jeu de données.

Parmi plusieurs choix de définition d'entropie possible la plus utilisée en traitement d'images est celle de Gull et Skilling (Gull and Skilling, 1991) :

$$H_g(S) = \sum_x \sum_y \left[S(x, y) - m(x, y) - S(x, y) \ln \left(\frac{S(x, y)}{m(x, y)} \right) \right] \quad (5.14)$$

où m est un modèle choisi pour être le fond du ciel et S le signal reconstruit. Cependant cette définition ne permet pas de valeurs négatives dans la solution, ce qui pose problème dans notre cas car les données expérimentales d'observations SZ

présentent de telles valeurs : on mesure des variations relatives de température. Afin de contourner ce problème il a été proposé Maisinger et al. (2004) de remplacer H_g par :

$$H_{+/-}(S) = \sum_x \sum_y \left[\psi(x, y) - 2m - S(x, y) \ln \left(\frac{\psi(x, y) + S(x, y)}{2m} \right) \right] \quad (5.15)$$

où $\psi(x, y) = \sqrt{S^2(x, y) + 4m^2}$. Ici m , à la différence de (5.14), est une constante fixée à la valeur rms attendue du signal.

Afin de résoudre les difficultés rencontrées par ces méthodes MEM pour traiter les images contenant à la fois des signaux à grandes et petites échelles angulaires, Pantin and Starck (1996) ont suggéré une nouvelle définition de l'entropie dans un cadre multi-échelles, décrit dans la section suivante. Il a alors été montré que les inconvénients des méthodes MEM (solutions dépendantes des modèles, sur-lissage des objets compacts, etc.) disparaissent.

5.3.2.2 Multiscale Entropy

La méthode d'entropie multi-échelles est basée sur les *a priori* standards des méthodes MEM et sur la décomposition en ondelettes du signal. L'idée est de considérer l'entropie du signal comme la somme des informations à chaque échelle de la décomposition. On utilise dans ce filtrage une décomposition multi-échelles UIWT, *Undeimated Isotropic Wavelet Transform*, basée sur un algorithme à trous avec une bspline cubique comme fonction d'échelle. La décomposition agit de la manière suivante sur une image S de taille $n \times n$:

$$S(x, y) = c(x, y) + \sum_{j=1}^J w_j(x, y) \quad (5.16)$$

où c est une version lissée de la carte originale S et les w_j représentent les détails de S présents à l'échelle 2^{-j} (Starck et al., 1998; Starck and Murtagh, 2002). À partir de cette décomposition et en notant $H(S)$ l'information relative au signal et $h(w_j(x, y))$ l'information relative à un seul coefficient d'ondelette, l'entropie multi-échelles est définie par :

$$H(S) = \sum_{x,y=1}^{N_j} h(c(x, y)) + \sum_{j=1}^{N_l} \sum_{x,y=1}^{N_j} h(w_j(x, y)) \quad (5.17)$$

où N_l est le nombre total d'échelles et N_j la taille de la carte à l'échelle j considérée. Dans cette relation on suppose l'indépendance des coefficients $w_j(x, y)$: l'entropie globale est alors la somme des entropies. Comme nous l'avons vu cette hypothèse n'est pas rigoureusement vraie, même pour les coefficients d'ondelette.

La fonction h dans l'équation 5.17 quantifie l'information portée par le coefficient d'ondelette $w_j(x, y)$. Plusieurs définitions ont été proposées pour la fonction h . Une discussion et comparaison des différentes définitions peut-être trouvée dans Starck

et al. (2005). La fonction h (l'entropie) est donnée, dans notre cas, en supposant un modèle de bruit, par l'équation suivante (voir Starck et al. (2001) pour tous les détails) :

$$h(w_j(x, y)) = \int_0^{|w_j(x, y)|} P_n(|w_j(x, y)| - u) \left(\frac{\partial h(t)}{\partial t} \right)_{t=u} du \quad (5.18)$$

où $P_n(w_j(x, y))$ est la probabilité que le coefficient $w_j(x, y)$ puisse être dû au bruit. Cette équation se simplifie dans l'hypothèse d'un bruit gaussien. Pour un bruit gaussien de déviation standard σ_j (dépendant de l'échelle j) on a :

$$\begin{aligned} P_n(x) &= \sqrt{\frac{2}{\pi}} \frac{1}{\sigma_j} \int_{|x|}^{+\infty} \exp(-W^2/2\sigma_j^2) dW \\ &= \operatorname{erfc} \left(\frac{|x|}{\sqrt{2}\sigma_j} \right) \end{aligned} \quad (5.19)$$

et 5.18 devient :

$$h(w_j(x, y)) = \frac{1}{\sigma_j^2} \int_0^{|w_j(x, y)|} u \operatorname{erfc} \left(\frac{|w_j(x, y)| - u}{\sqrt{2}\sigma_j} \right) du \quad (5.20)$$

Le modèle de bruit utilisé est le seul *a priori* de la méthode.

La composante SZ reconstruite après la séparation en composantes indépendantes \tilde{S}_{SZ} est bruitée. La méthode de filtrage présentée ici suppose que ce signal peut être décomposé comme suit :

$$\tilde{S}_{SZ} = S + N \quad (5.21)$$

avec S la carte contenant le signal physique et N un bruit additif. De la même manière on peut décomposer l'information contenue dans l'image SZ reconstruite en deux parties : la première H_s contenant l'information "utile", à savoir le signal SZ, et la seconde H_n contenant l'information "contaminante", à savoir le bruit :

$$H(\tilde{S}_{SZ}(x, y)) = H_s(\tilde{S}_{SZ}(x, y)) + H_n(\tilde{S}_{SZ}(x, y)) \quad (5.22)$$

Pour chaque coefficient d'ondelette $w_j(x, y)$ nous devons estimer les fractions h_s et h_n de h :

$$H(\tilde{S}_{SZ}(x, y)) = \sum_{j=1}^{N_l} \sum_{x, y=1}^{N_j} h_s(w_j(x, y)) + \sum_{j=1}^{N_l} \sum_{x, y=1}^{N_j} h_n(w_j(x, y)) \quad (5.23)$$

$$(5.24)$$

5.3.2.3 Filtrage ME-FDR

On peut formuler le problème de filtrage d'une carte bruitée S_n de la manière suivante. On cherche une carte filtrée S_f telle que :

1. la différence entre S_f et S_n présente un minimum d'information due au signal,
2. la carte S_f présente un minimum d'information due au bruit.

Ces deux “objectifs” sont en partie opposés. Un choix est nécessaire : d'un côté on veut éliminer un maximum de bruit (par un filtrage lourd) et de l'autre on veut retrouver un maximum de signal (par un filtrage léger). En pratique on minimise, au regard de $\tilde{w}_j(x, y)$, pour chaque échelle j la fonction suivante :

$$h_s(w_j(x, y) - \tilde{w}_j(x, y)) + \beta h_n(\tilde{w}_j(x, y)) \quad (5.25)$$

où $w_j(x, y)$ sont les coefficients d'ondelettes de la carte observée S_{obs} , $\tilde{w}_j(x, y)$ les coefficients d'ondelettes de la carte filtrée et β est le paramètre de régularisation quantifiant le compromis entre filtrage lourd ou léger.

On cherche à reconstruire complètement le signal des structures importantes (sans les filtrer de manière trop importante) tout en éliminant le bruit sur les petites structures. L'introduction d'un “support” multi-échelles Murtagh et al. (1995); Pantin and Starck (1996) va permettre de réaliser ceci. L'idée est d'appliquer le filtrage uniquement sur les coefficients d'ondelettes bruités, les autres composants de la cartes étant inchangés. La nouvelle définition de l'entropie multi-échelles devient :

$$\tilde{h}(w_j(x, y)) = M(j, k, l)h(w_j(x, y)) \quad (5.26)$$

où $M(j, k, l)$ est le “support” multi-échelles définit comme :

$$M(j, k, l) = \begin{cases} 0 & \text{si } w_j(x, y) \text{ est significatif} \\ 1 & \text{si } w_j(x, y) \text{ n'est pas significatif} \end{cases} \quad (5.27)$$

M indique si le coefficient $w_j(x, y)$ contient de l'information utile à une échelle j et une position (x, y) donnée. $w_j(x, y)$ est marqué comme significatif si la probabilité que le coefficient soit dû au bruit est faible, c'est à dire que ce coefficient est dû au signal cherché : il ne faut pas le filtrer. Inversement, $w_j(x, y)$ est marqué comme non significatif si la probabilité que le coefficient soit dû au bruit est forte, il faut alors appliquer le filtrage. Dans le cas d'un bruit gaussien, un coefficient $w_j(x, y)$ est significatif si $|w_j(x, y)| > k\sigma_j$ où σ_j est la déviation standard du bruit à l'échelle j , estimée à partir des données, et k est une constante. Sans un critère objectif pour choisir le seuil, il est choisi arbitrairement, généralement entre 3 et 5 (Murtagh et al., 1995).

La méthode FDR (*False Discovery Rate*), est un outil statistique dû à Benjamini and Hochberg (1995) qui offre une manière efficace de choisir un seuil adaptatif pour calculer le support multi-échelles. Cette technique a été décrite récemment par Miller et al. (2001); Hopkins et al. (2002); Starck et al. (2005) à travers différents exemples d'applications en astrophysique.

La méthode FDR fournit un moyen de contrôler de manière adaptative la fraction de fausses détections sur le nombre total de détections, c'est le rapport :

$$FDR = \frac{N_{Fausses}^{Detections}}{N^{Detections}} \quad (5.28)$$

où $N_{Fausses}^{Detections}$ est le nombre de pixels non-signifiactifs déclarés significatifs (pixels du fond déclarés amas), et $N^{Detections}$ est le nombre total de pixel détectés.

La méthode FDR assure, en adaptant la valeur de k , que ce rapport est en moyenne inférieur à α choisi entre 0 et 1. Nous choisissons $\alpha = 0.005$. Une description complète de la méthode FDR peut être trouvée dans Miller et al. (2001). Les articles de Hopkins et al. (2002) et de Starck et al. (2005) ont montré que cette méthode donne de meilleurs résultats que les méthodes standards de détection de sources. Nous utilisons la méthode FDR dans un cadre "multi-échelles" (voir Starck et al., 2005) : on choisit un seuil par échelle et un coefficient $w_j(x, y)$ est considéré comme significatif si sa valeur absolue est supérieure au seuil à cette échelle j .

En résumé. En supposant un bruit gaussien, la méthode ME-FDR cherche l'image qui minimise la fonction $U(S_f)$ considérant la carte bruitée S_n (voir l'équation 5.13) :

$$U(S_f) = \frac{\|S_n - S_f\|^2}{2\sigma_n^2} + \beta \sum_{j=1}^{N_l} \sum_{x,y} \tilde{h}_n((\mathcal{W}S_f(x, y))_j) \quad (5.29)$$

où σ_n est la déviation standard du bruit dans la carte observée S_n , N_l est le nombre d'échelles de la décomposition de la carte, \mathcal{W} est l'opérateur de décomposition en ondelettes. $\tilde{h}_n(w_{j,k,l})$ est l'entropie multi-échelles, non-nulle pour les pixels contenant du bruit, c'est à dire sur le support sélectionné par la méthode FDR. Des détails sur l'algorithme de minimisation peuvent être trouvés dans Starck et al. (2001). Dans notre simulation nous avons considéré un bruit gaussien uniforme sur toute la carte. La méthode est adaptable à un bruit gaussien non-uniforme.

5.3.3 Résultats

Les figures 5.3.1, 5.3.2 et 5.3.3 comparent les résultats des différents filtrages mis en oeuvre. Tous les filtrages sont appliqués directement sur la carte bruitée \tilde{S}_{SZ} obtenue après la séparation en composantes indépendantes par JADE travaillant à partir des données pré-filtrées et décomposées dans la base multi-échelles des ondelettes de Mallat. Les cartes de la figure 5.3.1 donnent un aperçu visuel du résultat de différents filtrages.

Filtrage gaussien. Nous comparons le filtrage ME-FDR à un filtrage Gaussien. Nous choisissons la taille du noyau gaussien comme étant la taille moyenne des lobes instrumentaux aux fréquences d'observation : $FWHM = 2.35$ arcmin. On teste par ailleurs avec les noyaux gaussien : $FWHM = 2$ arcmin et $FWHM = 3$ arcmin. On convolue la carte \tilde{S}_{SZ} par ces filtres. Les résultats pour les différents filtrages gaussiens utilisés sont donnés dans la figure 5.3.2.

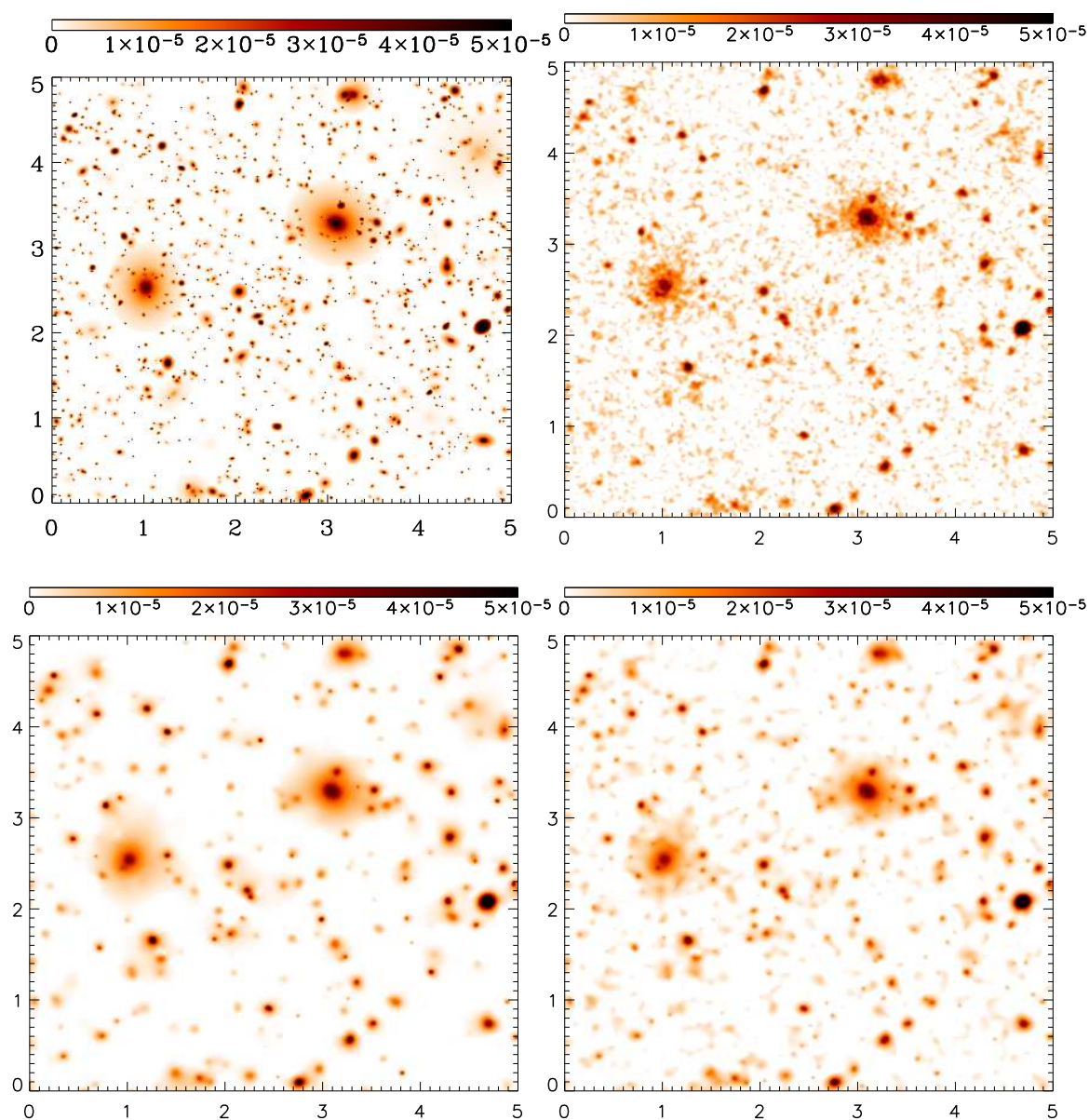


FIG. 5.3.1 – Cartes d’effet SZ : simulée (en haut à gauche) et reconstruite par JADE puis filtrées avec un filtrage gaussien (en haut à droite), ME-FDR seul (en bas à gauche) et avec un filtrage ME-FDR avec un pré-filtrage gaussien (en bas à droite).

1. La comparaison avec la courbe “*Sans filtrage*” (correspondant à la meilleure courbe de la figure 5.2.3 : “*Prefiltrage WJADE*”) permet de voir qu’aux petites échelles on élimine le bruit présent après la séparation en composantes indépendantes.
2. Aux grandes échelles le gain est négligeable : d’une part le bruit instrumental à ces échelles est inexistant d’autre part un noyau gaussien de la taille utilisée a peu d’influence sur ces échelles.

ME-FDR. Dans un premier temps nous appliquons le filtrage ME-FDR directement sur la carte bruitée, \hat{S}_{SZ} . Il est possible de choisir les échelles sur lesquelles faire travailler le filtrage ME-FDR. Une approche naturelle est d’éliminer l’échelle $l = 0$ ne contenant, a priori, que du bruit. Nous avons testé plusieurs options consistant à considérer toutes les échelles (ME-FDRF1), à enlever l’échelle 0 seulement (ME-FDRF2) et à enlever les échelles 0 et 1 (ME-FDRF3). Les résultats de ces filtrages montrent (figure 5.3.2) que :

1. plus de bruit est éliminé aux petites échelles,
2. aux grandes échelles le filtrage *introduit* des artefacts conduisant à une forte distorsion du signal (même les données non filtrées sont meilleures).

Il semble que le filtrage ME-FDR introduise artificiellement aux grandes échelles une distorsion. Le filtrage non-linéaire a pour but de “concentrer” l’énergie des structures afin d’éviter la dilution produite par un filtrage gaussien. Cette concentration agit aussi à grande échelle ce qui a pour effet de renforcer la composante de grande taille des structures mais aussi d’éventuels contaminants résiduels : tel que le CMB ou les poussières.

Afin d’améliorer le comportement du filtrage ME-FDR à grande échelle nous avons décidé de pré-filtrer les données par le noyau gaussien avec $FWHM = 2.35$ arcmin. Cette approche se justifie car nous avons fortement sur-échantillonné nos données et on ne perd, a priori, pas d’information en filtrant à la taille du lobe instrumental. Les résultats pour les différents filtrages ME-FDR utilisés sont donnés dans la figure 5.3.2. On constate sur cette figure que :

1. le préfiltrage corrige le problème du ME-FDR aux grandes échelles,
2. dégrade légèrement les performances du ME-FDR aux petites échelles,
3. les performances sont, à toutes les échelles, meilleures qu’un simple filtrage gaussien.

Conclusion La figure 5.3.3 compare les meilleurs filtrages gaussien et ME-FDR. Le filtrage ME-FDR précédé d’un pré-filtrage gaussien donne en moyenne le meilleur résultat sans introduire d’artefact aux grandes échelles. C’est le filtrage que nous utiliserons dans le chapitre 6.

Nous allons mettre en place dans la section suivante une méthode d’estimation de la qualité des différents filtrages en nous intéressant, non plus au comportement statistique des cartes d’erreur (c’est à dire à la distribution de l’intensité des pixels à chaque échelle), mais en étudiant, en vis-à-vis du catalogue d’amas simulé, le catalogue des sources brillantes détectées dans la carte reconstruite. Ce critère s’avère

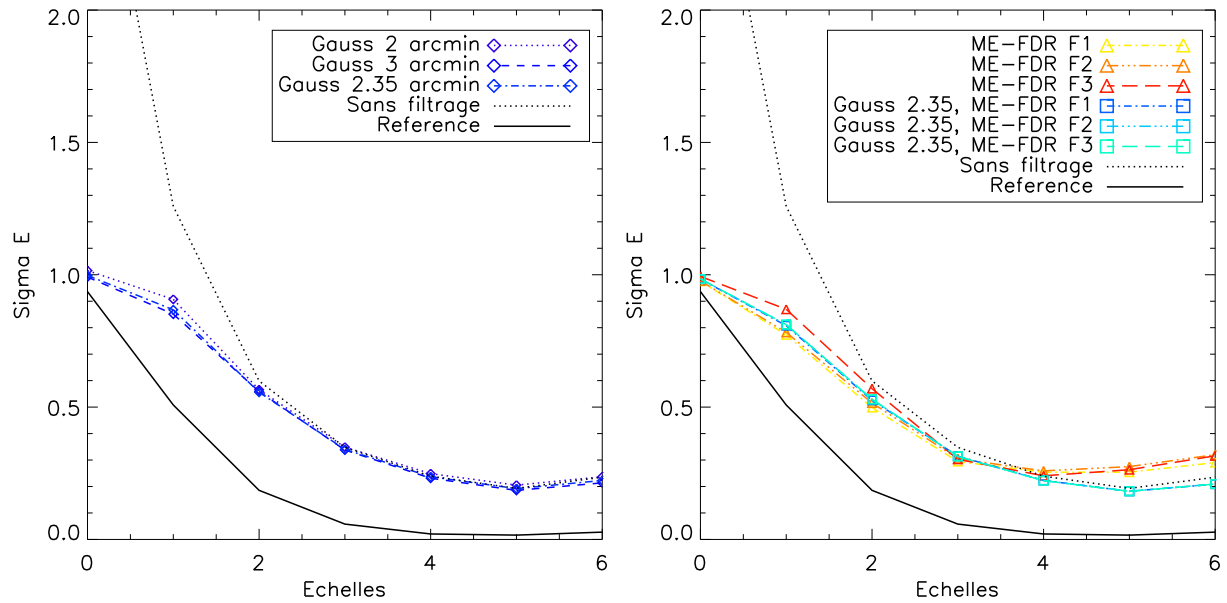


FIG. 5.3.2 – Résultats en termes de distorsion par échelle (voir le paragraphe 5.2.4.1) des filtres gaussien (à gauche) et ME-FDR (à droite).

beaucoup plus efficace et riche pour comparer la qualité des différents filtres. En effet le critère de distorsion utilisé dans cette section ne rend pas compte du caractère *faux* ou *vrai* d'un ensemble compact de pixels brillants. A l'aide des catalogues nous allons pouvoir distinguer les "fausses" sur-brillances (bruit, contamination, etc.) des "vraies" (les amas) et donc introduire de manière plus précise la notion de contamination et de *pureté* des données reconstruites et filtrées.

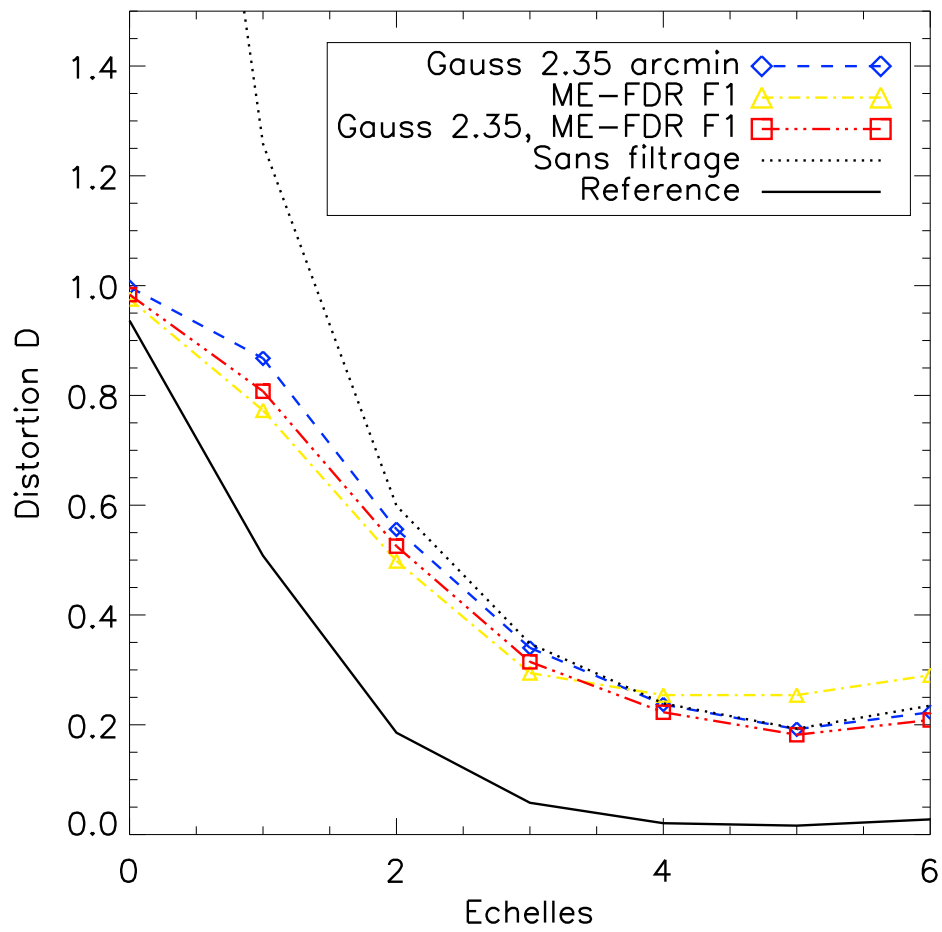


FIG. 5.3.3 – Comparaison des filtrages gaussien, ME-FDR et Gauss, ME-FDR.

5.4 Détection des amas

Les algorithmes ICA et de filtrage utilisés ont permis d'aboutir à une carte supposée ne contenir que des amas de galaxies (ICA) et débruitée (filtrage). Les cartes de la figure 5.4.1 montrent ces résultats. À partir de cette carte reconstruite on peut extraire un catalogue de sources brillantes (5.4.1), puis les classer en tant qu'amas de galaxies ou en tant que contamination (5.4.2). Enfin on peut quantifier les performances de la chaîne de détection en termes de complétude et de pureté de reconstruction du catalogue simulé d'amas (5.4.3).

5.4.1 Algorithme de détection de sources brillantes

Pour extraire les sources brillantes de la carte SZ reconstruite nous utilisons un logiciel public, SExtractor, de détection de sources (Bertin and Arnouts, 1996). SExtractor s'avère rapide, facilement configurable et bien adapté à l'extraction d'amas avec des tailles apparentes variant de quelques pixels de la carte à presque 60 pixels. Nous l'utilisons en configuration "sans bruit" : le filtrage que nous avons appliqué est considéré comme suffisant. On utilise ce logiciel pour l'identification des sources et comme outil de séparation (*deblending*), c'est à dire pour sa capacité à distinguer deux sources presque superposées, ce qui s'avère très utile pour les amas proches et de grande masse ayant une grande taille apparente sur le ciel. On peut voir sur la figure 5.4.1 dans le panneau en bas à droite, à la position (3.2, 3.4) degrés, un exemple de séparation réussie. Le tableau 5.4.1 donne la valeur des principaux paramètres de SExtractor que nous avons utilisés.

DEBLEND_MINCONT	0
DETECT_MINAREA	5
THRESH_TYPE	ABSOLUTE
DETECT_THRESH	seuil

TAB. 5.1 – Paramètres SExtractor utilisés pour l'extraction des amas dans la carte filtrée. Consulter Bertin and Arnouts (1996) pour plus de détails sur les paramètres et le fonctionnement du logiciel.

DEBLEND_MINCONT donne le contraste minimum pour séparer deux sources, une valeur de 0 donne une meilleure séparation des objets. DETECT_MINAREA donne la taille minimum des détections. Les amas lissés ont une taille minimum de 9 pixels : nous voulons être inclusif à cette étape de la détection. THRESH_TYPE : ABSOLUTE indique que le seuil de détection que nous utilisons est donné de manière absolue par rapport à zéro. Je reviendrai sur la partie "photométrie" de SExtractor dans le chapitre 6. DETECT_THRESH indique le seuil appliqué pour extraire les sources : SExtractor ne regarde, pour la détection, que les pixels plus brillants que ce seuil.

Ainsi SExtractor extrait, à partir de la carte reconstruite, un catalogue de sources contenant les informations suivantes : position dans la carte, taille, flux intégré, flux du pic central, paramètres de forme (ellipticité, orientation). Ce catalogue peut être

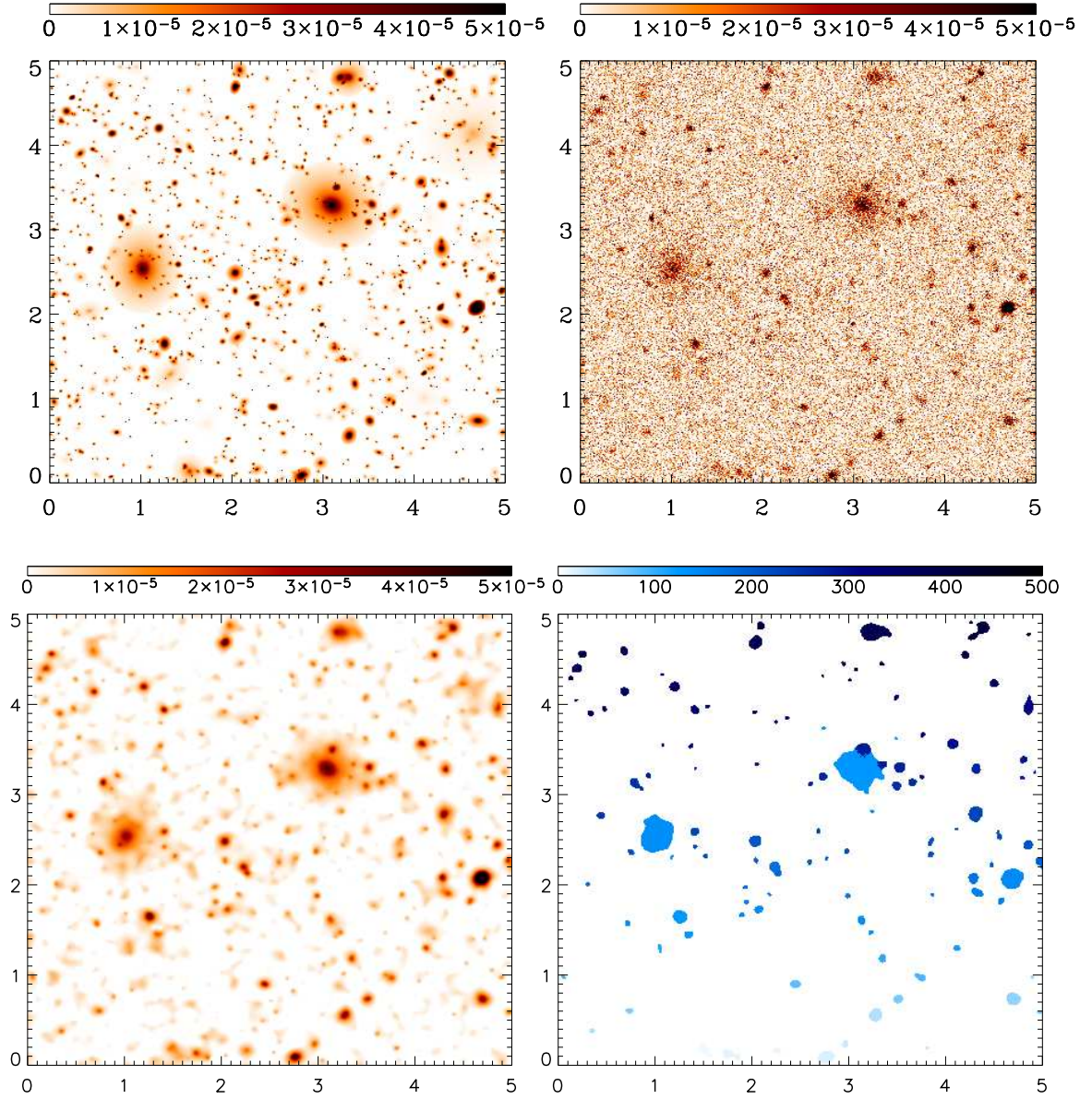


FIG. 5.4.1 – Cartes SZ aux différentes étapes de la chaîne de simulation-détection : simulée (en haut à gauche), reconstruite par JADE (en haut à droite), filtrée (en bas à gauche), et avec les sources détectées (en bas à droite).

contaminé, notamment lorsque le seuil de détection utilisé est bas. La contamination peut provenir du bruit instrumental non filtré ou d'une source infrarouge mal séparée par la méthode ICA. Dans le paragraphe suivant je présente un moyen de distinguer les vraies détections (des amas) des *fausses* détections (des contaminants).

5.4.2 Association et marquage des sources

Afin de classer les détections du catalogue reconstruit entre vraies et fausses détections on utilise un critère de distance (spatiale) par rapport au catalogue simulé d'amas. La démarche utilisée est la suivante :

1. on trie les détections par flux décroissant,
2. on fixe une distance seuil,
3. pour chaque détection, si un amas simulé se trouve à une distance inférieure à la distance seuil alors il est considéré comme "voisin" de la détection. La détection est alors marquée comme étant un amas. Si la détection ne possède pas de voisin dans le catalogue simulé d'amas alors elle est marquée comme une contamination.
4. Si une détection possède plusieurs voisins, le voisin ayant le flux intégré le plus grand est choisi comme amas simulé associé à la détection. Nous aurions pu choisir d'associer une détection au voisin le plus proche spatialement cependant nous avons constaté plus d'erreurs d'association par cette méthode.

Ces deux critères (distance spatiale seuil et amas le plus brillant) permettent de classer les détections comme amas ou contaminants. Le critère de distance spatiale nécessite de prendre des précautions. La figure 5.4.2 permet de comprendre le problème. Pour un seuil de détection y_δ bas on distingue clairement deux populations de détections. Les amas ont leur plus proche voisin simulé à une distance faible (confirmé par la courbe à seuil de détection haut où l'on a alors que des amas), généralement inférieure à quatre pixels. Les fausses détections ont leur plus proche voisin simulé à une distance généralement supérieure à cinq pixels : c'est la partie plate de la courbe à grande distance. Le critère de distance s'applique alors de manière naturelle. Le seuil introduit ici est un seuil sur la valeur des pixels, c'est à dire un seuil en paramètre de Compton sur la ligne de visée : y_{seuil} . Dans le cas où le seuil de détection est trop bas : les deux distributions de "distance au plus proche voisin simulé" se superposent. Il est alors difficile d'utiliser le critère de distance tel qu'il a été défini car l'erreur sur le marquage devient trop importante. En pratique on se place toujours dans le cas où il est possible de distinguer les deux distributions ou dans le cas où la distribution des contaminations est faible par rapport à la distribution des vraies détections. Les distributions des distances à seuil élevé nous guide pour choisir la distance limite à 4 pixels. Le seuil le plus bas applicable pour une chaîne de détection donnée est déterminé par l'examen des distributions de distance pour les différents seuils de détections y_δ . A 4 pixels de distance on veut moins de 2% de fausses détections vis-à-vis du pic à 1 pixel.

- ME-FDR F1 : on choisit le seuil limite à $y_{min} = 3 \times 10^{-6}$
- Gaussien 2.35 arcmin : on choisit le seuil limite à $y_{min} = 8 \times 10^{-6}$
- Gaussien 2.35 arcmin, ME-FDR F1 : on choisit le seuil limite à $y_{min} = 5 \times 10^{-6}$

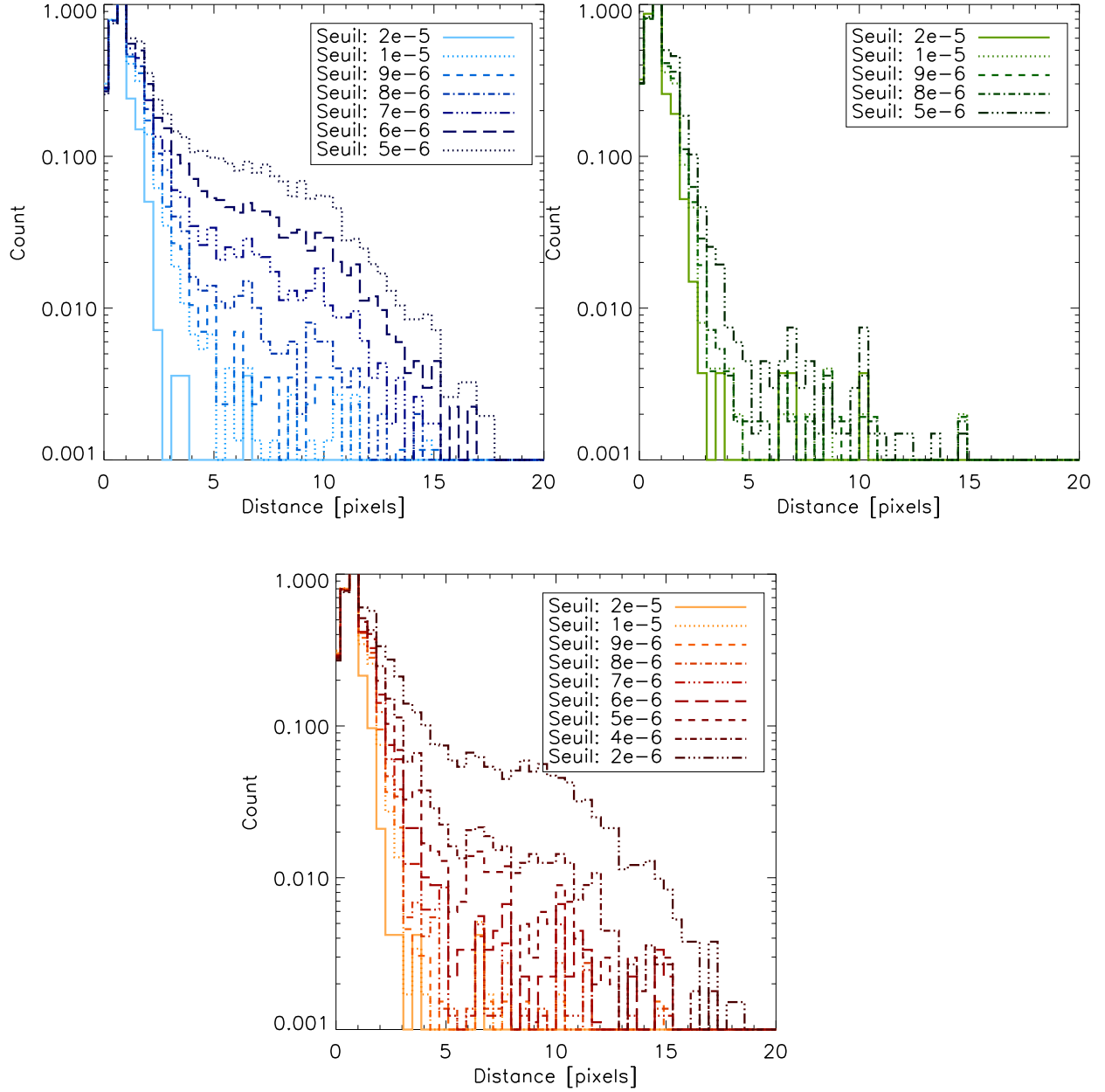


FIG. 5.4.2 – Histogrammes des distances entre sources détectées et amas simulés pour différents filtrages et différents seuils de détection y_δ ; en haut à gauche pour le filtrage gaussien, en haut à droite pour le filtrage ME-FDR et en bas pour le filtrage Gauss, ME-FDR. Ces histogrammes permettent de fixer la distance limite de marquage des sources.

Appliquer un seuil plus bas que ce seuil limite introduirait des erreurs de marquage.

Lorsqu'on baisse le seuil de détection le risque de mauvais marquages augmente. Pour le moment nous nous sommes placés dans des cas où cette erreur est négligeable. Une quantification précise pourra être envisagée ultérieurement.

5.4.3 Qualité du catalogue reconstruit

Les critères de marquage introduits au paragraphe précédent permettent de qualifier la qualité des différentes chaînes de détections mises en place (JADE, JADE avec ondelettes, filtrage gaussien, filtrage linéaire, etc.) en vue de l'objectif que nous nous sommes fixé : détecter les amas en tant qu'objets. Nous définissons deux quantités permettant de décrire la qualité de ces méthodes :

- la pureté (*purity*) :

$$\text{Purity}(y_\delta) = \frac{N_{det}^{true}(y_\delta)}{N_{det}(y_\delta)} \quad (5.30)$$

avec N_{det}^{true} les vraies détections (amas) et N_{det} toutes les détections (amas et contamination) trouvées pour un seuil de détection y_δ donné. Cette quantité donne le taux de pureté des détections.

- La complétude (*completeness*) :

$$\text{Completeness}(y_\delta) = \frac{N_{det}^{true}(y_\delta)}{N_{simu}} \quad (5.31)$$

est aussi fonction du seuil de détection.

Nous nous intéressons dans ce chapitre à la comparaison de chaînes de détection d'amas. Nous remettons au chapitre 6 le travail de qualification de la photométrie des détections. La quantité N_{simu} définie précédemment est choisie indépendante du seuil : c'est le nombre total d'amas dans la simulation. Cette définition est particulière mais elle permet de comparer facilement les différentes méthodes introduites ici. Nous utiliserons dans le chapitre suivant la définition générale de la complétude et de la pureté en fonction du seuil en paramètre de Compton *intégré* : Y_{seuil} .

5.4.4 Résultats

Nous présentons les résultats de l'application de ces critères de qualité dans la figure 5.4.3. Les premières figures donnent la pureté et la complétude du catalogue reconstruit en fonction du seuil de détection appliqué. Le seuil minimum est fixé, en accord avec le paragraphe précédent, de manière à ce que l'erreur de marquage soit négligeable. La seconde figure donne la complétude du catalogue reconstruit en fonction de la pureté de celui-ci. A une pureté donnée, soit un seuil de détection donné, le meilleur catalogue est celui qui fournit la plus grande complétude. Cette figure permet de comparer la qualité des méthodes.

On constate ainsi que :

1. à bas seuil, le filtrage gaussien comporte beaucoup de contaminations : la pureté chute rapidement, tandis que le filtrage ME-FDR F1 garde une excellente pureté,

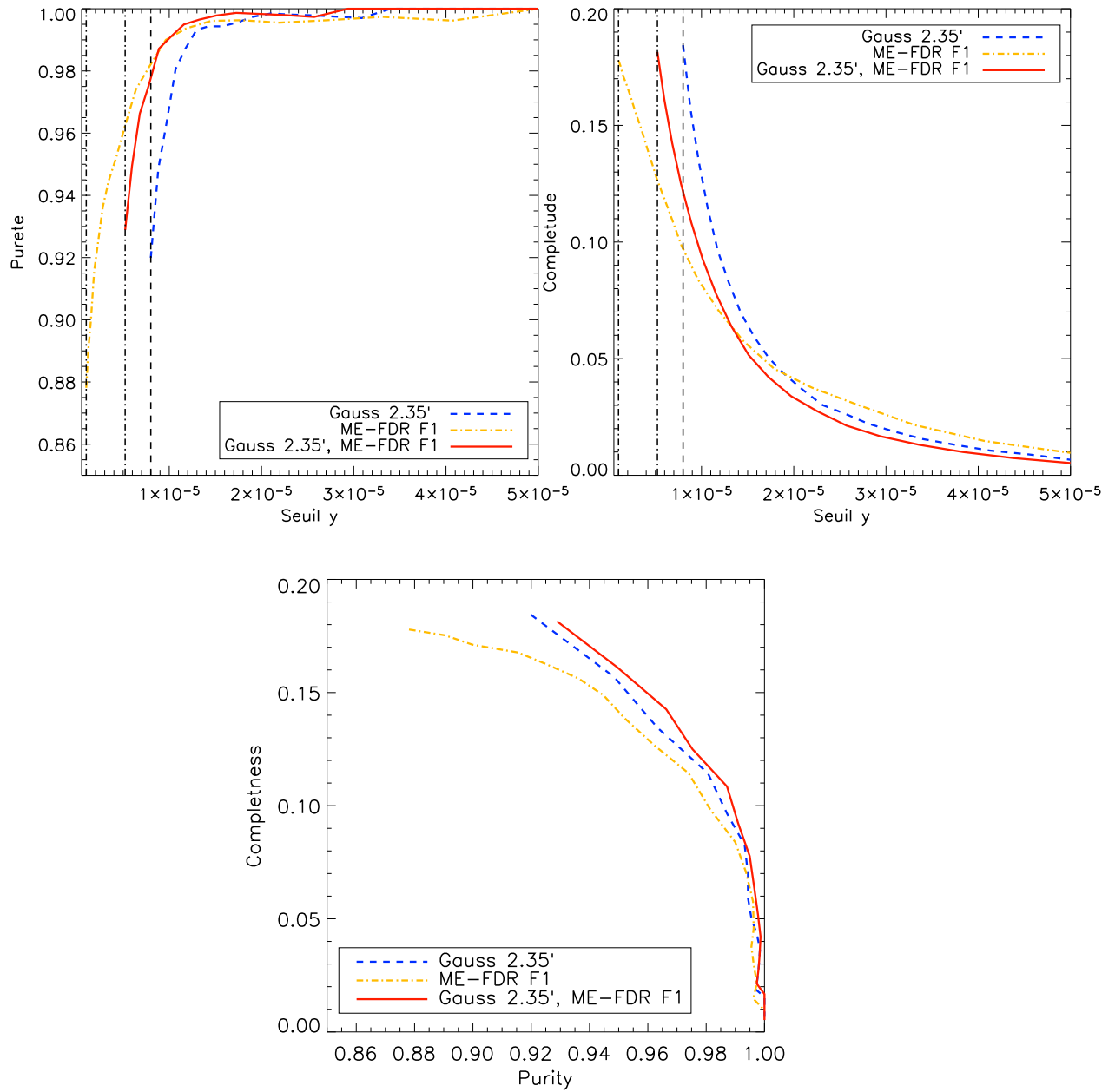


FIG. 5.4.3 – Comparaison de la qualité des différents filtrages : en haut à gauche la pureté du catalogue de sources détectées en fonction du seuil de détection, en haut à droite la complétude en fonction du seuil et en bas la complétude en fonction de la pureté du catalogue reconstruit.

2. toujours à bas seuil, le filtrage gaussien fournit le plus grand nombre de détections (amas et contaminations), contre balançant la faible pureté ce qui conduit malgré tout à la complétude la plus haute. Le filtrage ME-FDR laisse, quant à lui, peu de détections à bas seuil.
3. A haut seuil, en revanche le filtrage ME-FDR a une mauvaise pureté. Ceci s'explique par le fait que ce filtrage a tendance à concentrer l'énergie des détections : vraies et fausses, ainsi des contaminations se retrouvent à haut seuil, ce qu'évite un pré-filtrage gaussien comme le montre la pureté du filtrage Gauss-ME-FDR.
4. Le filtrage Gauss-ME-FDR, qui est pourtant intermédiaire en terme de pureté et de complétude, se trouve avoir, quelle que soit la pureté du catalogue, la meilleure complétude. Ceci confirme les résultats précédents en termes de distorsion des cartes d'erreur (voir la section 5.3.3).

Je rappelle que la complétude calculée ici, est rapportée au nombre total d'amas dans la simulation.

5.5 Conclusion

Le coeur du premier élément de la chaîne de détection (un schéma résumant la méthode est donné dans la figure 5.5.1) est constitué du logiciel Jade utilisant un algorithme ICA *blind*. Le rôle de cette première étape est de séparer les différentes sources astrophysiques et d'isoler le signal SZ des autres signaux. Pour revenir à la discussion sur la simulation, le premier défi de cet algorithme est de pouvoir éliminer les galaxies infrarouges de la carte reconstruite “contenant” l'effet SZ.

La première partie de mon travail a été de mettre en place un critère de qualité pour évaluer les performances de la séparation en composantes indépendantes. Une estimation de la distorsion du signal recherché, par échelle spatiale, est un critère de qualité pertinent. Il permet, non seulement de classer les différents algorithmes testés, mais aussi de localiser à quelle échelle se situe la distorsion (quelle taille apparente typique). C'est un point important car les signaux que l'on cherche à séparer possèdent des tailles apparentes qui vont de quelques pixels pour les galaxies infrarouges, à plusieurs degrés pour la poussière ou le CMB. Les amas se situant entre les deux, on est ainsi en mesure de repérer l'origine de certaines contaminations ou sources d'erreur. Ce critère servira aussi, notamment, pour estimer la qualité des filtrages.

Les caractéristiques particulières du signal SZ (par rapport aux autres signaux), en terme de spectre électromagnétique et de cohérence spatiale, ont guidé notre choix vers une “préparation” du signal observé, avant l'utilisation de Jade. Les signaux observés sont décomposés en leurs composantes à différentes échelles, allant de moins d'une minute d'arc à plusieurs degrés. Cette approche permet de mieux décrire les différents signaux et aide Jade à mieux les séparer. Un autre choix s'est révélé pertinent : un pré-filtrage des données éliminant le bruit des petites échelles spatiales permet d'améliorer les performances de la séparation. Ces deux points, faisant toute l'originalité de la méthode, permettent de rendre la méthode robuste à un haut niveau de bruit. Ces deux choix ont permis de mettre en place le premier élément de la chaîne d'extraction d'amas SZ qui répond au “défi” de l'extraction des amas : la carte reconstruite SZ contient une part négligeable de signal issu des galaxies infrarouges, limitant la contamination du catalogue final (i.e. augmentant sa pureté). Un inconvénient, connu, de la méthode qu'il convient de souligner ici est la nécessité d'étalonner chacune des composantes reconstruite sur une source connue par ailleurs. Un gros travail est à fournir pour développer un algorithme ICA aveugle, permettant de traiter ce point. J'ai supposé, dans mon travail, pouvoir accéder à une telle source, constituée de cent amas brillants, observés par ailleurs.

Le second élément de la chaîne d'extraction d'amas SZ est constitué d'un algorithme de filtrage non-linéaire utilisant une méthode de maximisation d'entropie : ME-FDR. Le rôle de cette étape est de maximiser le signal-sur-bruit afin de limiter, encore, la contamination présente dans le catalogue reconstruit. Par ailleurs, afin de remplir le critère photométrique, le filtrage ne doit pas trop affecter le signal.

Le filtrage ME-FDR se révèle tout a fait adapté pour le traitement de la carte SZ reconstruite, isolée lors de la première étape de la chaîne. En effet, par l'utilisation d'un support et d'un *a priori* entropique, ce filtrage adapte son “intensité” en fonction, non seulement de la position spatiale considérée mais aussi à l'échelle considérée. Les zones contenant du bruit sont fortement filtrées, tandis que les zones contenant

déjà un haut signal-sur-bruit, c'est à dire les amas brillants, sortant du bruit, ne sont que très peu affectées.

Ce filtrage permet deux choses : obtenir une contamination du catalogue final raisonnable et limiter l'impact du débruitage sur la photométrie des sources détectées. La carte SZ reconstruite se révèle être très fidèle à l'originale (voir la figure 5.4.1).

Le dernier élément de la chaîne d'extraction d'amas est constitué du logiciel public SExtractor, écrit par Emanuel Bertin. Son rôle dans notre chaîne d'analyse est d'identifier, au dessus d'un seuil de détection, les sources brillantes, de les cataloguer et d'en faire la photométrie.

Mon choix s'est porté sur ce logiciel car il présente un algorithme de *deblending* permettant de traiter, au moins en partie, les cas de confusion des sources, importante dans le cas des amas de galaxies dont la taille apparente varie énormément et atteint, parfois, le degré. Le second point ayant conduit au choix de ce logiciel est le fait qu'il propose un algorithme de photométrie donnant le flux intégré d'une détection : il s'adapte à sa taille. Ce dernier point, associé au fait que cette photométrie est faite sans connaissance *a priori* sur le profil des amas, et au fait que le filtrage ME-FDR n'agit que peu sur les sources brillantes, font de SExtractor un outil bien adapté à la chaîne d'analyse mise en place.

L'ellipticité des amas détectés par SExtractor n'est pas directement reliée à l'ellipticité simulée des amas. Il y a plusieurs raisons à cette constatation : 1) le lobe instrumental, simulé symétrique, a tendance à arrondir les structures, 2) le filtrage ME-FDR, utilisant une décomposition en ondelettes isotropiques, a aussi tendance à arrondir les structures, 3) la présence de bruit instrumental et les empilements d'amas rendent la mesure de l'ellipticité "vraie" impossible avec la chaîne de détection présentée ici. Il serait intéressant d'estimer l'erreur induite par cette hypothèse d'amas elliptique sur, par exemple, les erreurs photométriques présentées au chapitre 6. SExtractor est configuré pour calculer un flux intégré et il tient compte de l'ellipticité des structures.

Grâce au catalogue reconstruit nous avons pu estimer de manière plus précise la qualité des différentes chaînes d'analyse mises en place. Le calcul de la complétude en fonction de la pureté pour les différentes méthodes a montré (voir la figure 5.4.3) que la méthode la plus performante est la suivante :

1. pré-filtrage des cartes observées à la taille du lobe d'observation instrumental,
2. transformation du signal (cartes observées lissées) dans la représentation multi-résolution de Mallat,
3. application de l'algorithme ICA JADE sur ces données : estimation de la matrice de mélange,
4. utilisation de la matrice de mélange estimée sur les données bruitées afin d'isoler au mieux le signal SZ,
5. pré-filtrage des cartes par un noyau gaussien de largeur adaptée au lobe instrumental,
6. filtrage des cartes par l'algorithme non-linéaire ME-FDR,
7. application de SExtractor à la carte filtrée pour extraire un catalogue d'amas.

En termes d'algorithme de marquage des sources on pourrait améliorer les choses de différentes manières :

- par la mise en place d'un choix automatique du seuil en utilisant une méthode FDR, assurant alors un taux de mauvais marquage plus bas qu'une limite fixée.
- Parmi les contaminations on pourrait envisager un marquage distinguant une galaxie infrarouge d'une autre contamination, en comparant le catalogue des contaminations au catalogue simulé des galaxies infrarouges avec le même genre de critère de distance et de flux.
- Nous avons utilisé le premier ordre de la notion de "distance photométrique" (distance entre deux objets faisant intervenir leur photométrie : flux total, taille). On pourrait envisager de définir une vraie distance photométrique. Une définition possible est la suivante :

$$d_{phot}^2 = r_{phot}^T \cdot C \cdot r_{phot}$$

où $r_{phot}^T = (x_{pos}, y_{pos}, Y, \theta)$ est le vecteur "position photométrique" de l'objet et C la matrice de variance-covariance des positions photométriques calculées à partir des catalogues marqués, au préalable, par une méthode plus simple (par exemple celle utilisée dans notre cas avec un seuil de détection haut). Cette distance devrait être plus discriminante que la distance uniquement spatiale et permettre un marquage plus précis.

Le travail présenté dans ce chapitre fait l'objet d'une publication (soumise à A&A) écrite en collaboration avec : S. Pires (CEA-Dapnia/SAP), D. Yvon (CEA-Dapnia/SPP), Y. Moudden (CEA-Dapnia/SAP), S. Anthoine () et E. Pierpaoli () : astro-ph/0508641. Dans cet article nous comparons les performances de notre méthode avec celle développée par S. Anthoine et E. Pierpaoli et avec l'utilisation d'un filtrage de Wiener en plus du filtrage gaussien.

Dans le chapitre suivant nous mettrons en place, à partir de cette chaîne de détection, la fonction de sélection d'Olimpo l'un des futurs instruments d'observation grand champ d'amas SZ (Olimpo, APEX, Planck). Elle nous permettra de mettre en place un modèle statistique d'instrument. Nous quantifierons les erreurs photométriques et l'efficacité de détections des amas (complétude) ainsi que la contamination (par les sources ponctuelles infra-rouges ou autre) présentent dans les cartes.

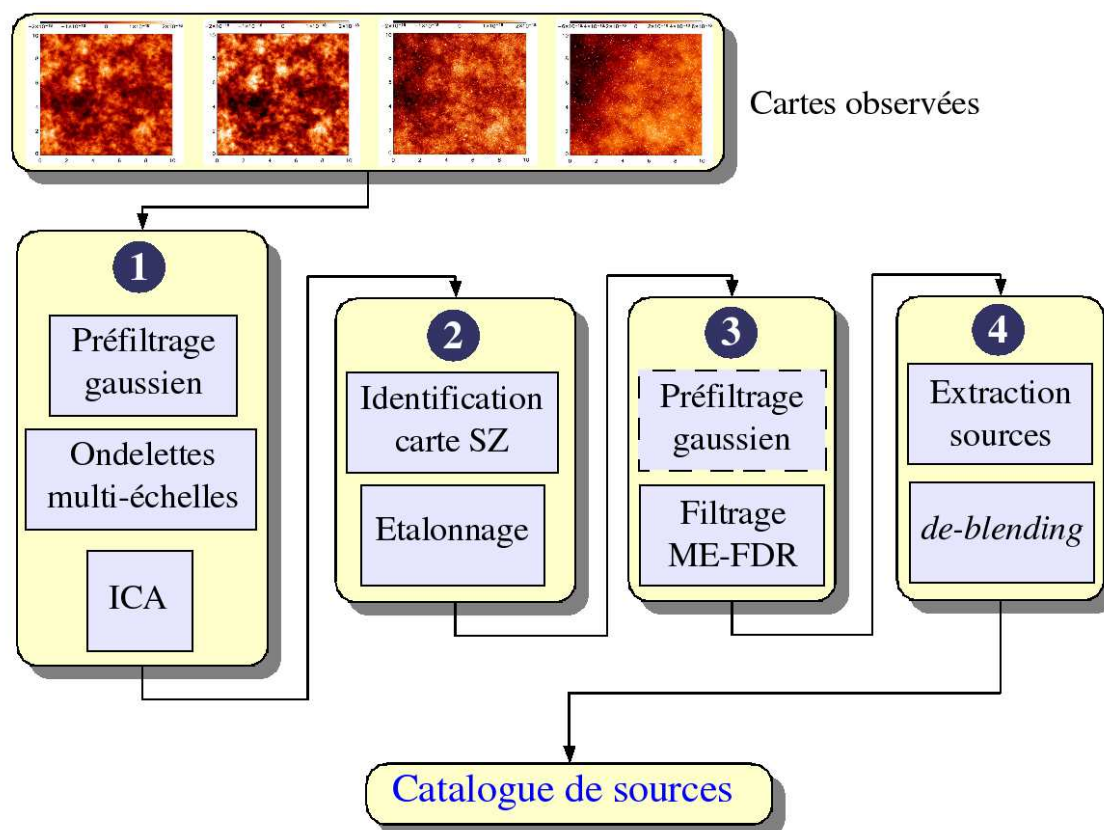


FIG. 5.5.1 – Schéma de la chaîne de détection d'amas SZ. On distingue quatre étapes importantes : 1) la séparation en composantes indépendantes, 2) la normalisation de la carte SZ estimée, 3) le filtrage de la carte bruitée, 4) l'identification des sources et la construction du catalogue. Cette chaîne, en simulation, s'accompagne du travail d'association des sources détectées avec le catalogue d'amas simulé.

Chapitre 6

Modélisation statistique des effets de sélection

Sommaire

6.1	Introduction : les effets de sélection	136
6.2	Modèle statistique d'Olimpo	139
6.2.1	Complétude	140
6.2.2	Erreurs photométriques	143
6.2.2.1	Estimation de Y_{obs}	144
6.2.2.2	Estimation de z_{obs}	149
6.2.3	Les contaminations	149
6.2.4	Nombre effectif de détections	151
6.3	Validité du modèle statistique	154
6.3.1	Complétude, photométrie et contaminations	154
6.3.2	Modèle statistique et confusion	157
6.4	Conclusion	163

6.1 Introduction : les effets de sélection

Je décris dans ce chapitre comment calculer les effets de sélection des futurs instruments d'observation grand champ SZ en utilisant la chaîne d'analyse présentée dans le chapitre précédent. Les effets de sélection, décrits par un modèle statistique, quantifient les performances de l'instrument et de la chaîne d'analyse des données. La connaissance du modèle statistique de la chaîne complète de détection, est essentielle pour pouvoir calculer correctement les contraintes apportées aux modèles cosmologiques ou aux modèles de formation de structure. En effet, ce modèle permet d'estimer le modèle cosmologique sous-jacent à partir d'un catalogue observé d'amas. Je reviendrai en détail sur ce point au chapitre 7.

Les effets de sélection ont plusieurs origines : astrophysiques, instrumentales ou algorithmiques, les trois étant étroitement liés. Les effets astrophysiques concernent la présence de contaminants (sources ponctuelles, résidus de CMB ou de poussière galactique) qui empêchent d'isoler correctement les amas. Les effets instrumentaux sont le résultat d'imperfections "normales" de l'instrument : résolution angulaire, bruit instrumental, observation à un nombre de fréquences limité, etc. Les effets algorithmiques proviennent de l'imperfection des outils d'analyse utilisés : déformation du signal, "bande passante" d'un filtrage trop étroite, trop large, etc. On voit ici que les trois effets sont liés : une méthode de séparation en composantes qui fonctionne bien (effet de sélection algorithmique) permet de s'affranchir, par exemple, de certaines sources de contamination astrophysique (effet de sélection astrophysique), et cette méthode fonctionne car les caractéristiques de l'instrument le permettent (effet de sélection instrumental). La fonction de sélection introduite dans la section 6.2 permet de décrire les performances de toute la chaîne de détection des amas par effet SZ : depuis l'instrument jusqu'à la chaîne d'analyse et de traitement du signal. Tout d'abord je rappelle la définition théorique de la fonction de sélection ainsi que, pour chacune de ses composantes, le moyen utilisé pour la calculer. Dans la section 6.3 je discute la validité du modèle statistique de la chaîne de détection utilisé.

A partir des calculs de la fonction de sélection nous allons pouvoir construire un modèle statistique de la "chaîne de détection SZ" complète depuis la prise des données par l'instrument jusqu'à leur analyse finale. Ce modèle d'observation est calculé dans un cadre cosmologique donné. La question que l'on peut alors se poser concerne la dépendance (ou l'indépendance) de cette fonction de sélection vis-à-vis du modèle cosmologique. Nous supposons que le modèle statistique de la *chaîne de détection* est *indépendant de la cosmologie* dans lequel il a été calculé. En pratique, quelle peut être l'influence de la cosmologie sur le modèle statistique ? Elle peut provenir :

- tout d'abord des amplitudes relatives des différentes composantes physiques présentes : les algorithmes d'analyse pourraient fonctionner moins bien dans des conditions différentes, par exemple, les anisotropies du CMB peuvent, dans certains modèles cosmologiques, être plus proche de la forme des amas (même échelle typique sur le ciel) que dans d'autres.
- Ensuite, au niveau des amas eux-mêmes, du bruit de confusion (superposition des sources) qui est une source importante de dégradation des performances d'une chaîne de détection SZ (voir la section 6.2.2). Cet effet est étroitement

dépendant de la taille du lobe instrumental et de l'algorithme traitant les phénomènes de superposition. Le modèle cosmologique sous jacent, parce qu'il implique une corrélation angulaire de la position des amas et la formation de la fameuse structure filamentaire des grandes structures, influence la confusion des sources. Un modèle cosmologique conduisant à une forte corrélation de la position des structures conduit à des cartes d'amas où la confusion des sources est importante et donne probablement un modèle statistique de chaîne de détection SZ différent que s'il est calculé dans un cadre cosmologique où les corrélations en position des amas sont faibles (peu de confusion des sources).

- Enfin la confusion des sources dépend de leur taille apparente moyenne. Certains modèles cosmologiques prédisent, pour un flux et un redshift donné, des masses et donc des tailles apparentes différentes de celles calculées dans, par exemple, le modèle de concordance.

Ces effets sont complexes et nous avons décidé de garder cette hypothèse de travail : nous nous plaçons dans un premier temps dans le cadre cosmologique du modèle de concordance. Nous irons plus loin en fin de chapitre et au chapitre 7.

La seule relation entre cosmologie et superposition des amas qui intervient dans notre simulation passe à travers la densité d'amas par unité de surface. Cette quantité dépend principalement du paramètre de formation de structures σ_8 . À défaut, d'une modélisation complète qui dépendrait de la cosmologie, une simulation Monte-Carlo systématique pour chaque modèle cosmologique testé permettrait de prendre en compte correctement tous ces effets de sélection.

Dans les chapitres 4 et 5 nous avons présenté une simulation d'observation de type Olimpo et une méthode d'extraction d'amas. Les résultats sont calculés dans le cadre du modèle cosmologique dont les paramètres sont donnés dans le tableau 4.1. Ce modèle, ayant une valeur de σ_8 importante, présente une grande densité surfacique d'amas. Pour les chapitres 6 et 7 nous nous plaçons dans le cadre du modèle cosmologique de concordance dont les paramètres sont donnés dans le tableau 6.1.

Ω_{tot}	Ω_b	Ω_Λ	Ω_{DM}	h	n_s	σ_8	f_{masse}	T_\star	f_g
1	0.04	0.7	0.26	0.7	1	0.85	S. & T.	1.6	$0.9 \Omega_b / \Omega_m$

TAB. 6.1 – Modèle cosmologique de concordance servant de base à la caractérisation de la chaîne de détection SZ mise en place dans ce chapitre.

L'étude présentée au chapitre 6 a été appliquée dans ce cadre et conduit au choix du filtrage ME-FDR F2, dont les performances en terme de complétude sont moindre que celle du filtrage ME-FDR F1 précédé d'un préfiltrage gaussien (on perd approximativement 150 détections sur 700), mais il donne, au seuil de détection choisi, une pureté plus grande pour le catalogue des sources extraites : on choisit ici de faire ce compromis afin de limiter la contamination du catalogue au maximum.

Dans ce chapitre je réponds aux questions suivantes :

- Compte tenu des caractéristiques de l'instrument, quels amas, en terme de flux, masse, redshift, taille, vais-je pouvoir observer et combien ?
- Dans quelle mesure vais-je pouvoir reconstruire leur flux ?

- Quelle contamination puis-je attendre sur le catalogue d'amas que je vais extraire? En d'autres termes : quelle est sa pureté?

6.2 Modèle statistique d'Olimpo

Je présente dans cette section les méthodes mises en place pour calculer le modèle statistique d'Olimpo. Ces méthodes sont facilement adaptables pour décrire d'autres expériences bolométriques multi-fréquences (type APEX, SPT, Planck, etc.). Le modèle statistique d'une chaîne de détection permet de calculer, à partir de la distribution théorique des amas de galaxies en fonction de leur flux et de leur redshift, la distribution effectivement observée par l'instrument et extraite par la chaîne d'analyse :

$$\frac{dN_{th}}{dz_{th}dY_{th}}(z_{th}, Y_{th}) \longrightarrow \text{Modèle statistique} \longrightarrow \frac{dN_{obs}}{dz_{obs}dY_{obs}}(z_{obs}, Y_{obs})$$

avec z_{obs} et Y_{obs} les observables effectivement mesurées des sources détectées dans l'expérience et z_{th} et Y_{th} les flux et redshift "vrais" de ces sources. Les observables utilisées pour l'expérience Olimpo sont le paramètre de Compton intégré des amas Y_{obs} et leur redshift z_{obs} . Dans la section 6.2.2 je montrerai que l'information sur la taille des amas est difficilement exploitable avec l'algorithme de photométrie mis en place ici.

La distribution observée des amas s'écrit, dans le cas le plus général, sous la forme suivante :

$$\begin{aligned} \frac{d\bar{N}_{obs}^{det}}{dz_{obs}dY_{obs}}(z_{obs}, Y_{obs}) = & \iint \left(\frac{d\bar{N}_{th}^{amas}}{dz_{th}dY_{th}}(z_{th}, Y_{th}) \times F(z_{th}, Y_{th}) \right) \\ & \cdot P(z_{obs}, Y_{obs} | z_{th}, Y_{th}) dY_{th} dz_{th} \\ & + \frac{d\bar{N}_{obs}^{cont}}{dz_{obs}dY_{obs}}(z_{obs}, Y_{obs}) \end{aligned} \quad (6.1)$$

avec :

- a) la **complétude** du sondage (déjà définie au chapitre précédent dans une version simplifiée) : $F(z_{th}, Y_{th})$,
- b) les **erreurs photométriques** sur les observables z_{obs} et Y_{obs} : $P(z_{obs}, Y_{obs} | z_{th}, Y_{th})$,
- c) les **contaminations** : $\frac{d\bar{N}_{obs}^{cont}}{dz_{obs}dY_{obs}}(z_{obs}, Y_{obs})$.

\bar{N}_{obs}^{det} est le nombre de détections extraites : $\bar{N}_{obs}^{det} = \bar{N}_{obs}^{amas} + \bar{N}_{obs}^{cont}$, \bar{N}_{th}^{amas} est le nombre théorique d'amas simulés dans la carte et \bar{N}_{obs}^{cont} est le nombre de détections identifiées comme étant des contaminations. \bar{N}_{obs}^{amas} est le nombre de détections identifiées comme étant des amas. Ces grandeurs sont des quantités moyennes. Afin de compléter le modèle statistique on construit aussi, par des simulations Monte-Carlo, les distributions du nombre effectif de détection, du nombre effectif d'amas détectés et du nombre effectif de contaminations obtenues dans une carte. On montre dans le paragraphe 6.2.4 que ces distributions ne sont pas poissonniennes. Dans les paragraphes suivants nous travaillons avec les valeurs moyennes.

Le calcul pratique pour obtenir le modèle statistique de chaque chaîne de détection SZ (instrument et analyse des données) s'effectue au moyen de N générations de la chaîne de simulation complète suivante :

1. Simulation des cartes du ciel : calcul du modèle cosmologique, calcul des cartes physiques du ciel aux fréquences d'observation.
2. Simulation de l'instrument : convolution des cartes par le lobe instrumental, ajout du bruit de mesure.
3. Séparation en composantes indépendantes : pré-filtrage des cartes observées, décomposition multi-échelles du signal, calcul de la matrice de mélange, application de cette matrice aux cartes observées bruitées.
4. Filtrage par l'algorithme ME-FDR F2.
5. Extraction des sources brillantes : utilisation de SExtractor et obtention d'un catalogue de détections.
6. Marquage des détections : on utilise l'algorithme de marquage des sources présenté dans la section 5.4.2, chaque détection est classée parmi les amas ou parmi les contaminations.

A partir des N générations on construit un catalogue des amas simulés et un catalogue des détections. Le seuil de détection y_δ pour l'extraction des sources est choisi de manière à obtenir un catalogue d'une pureté donnée (voir la section 5.4.3).

Dans le chapitre précédent les complétude et pureté présentées sont des grandeurs "cumulatives" : on regarde toutes les détections *au-dessus* d'un seuil et non pas *autour* de ce même seuil. Dans la fonction de sélection on s'intéresse aux grandeurs observables des amas et à la répartition des amas autour de ces grandeurs : on parle de complétude et de contamination "différentielles". En pratique le seuil de détection y_δ est choisi supérieur à y_{min} , qui vaut ici 7×10^{-6} afin d'éviter les erreurs de marquage. Dans le cas d'Olimpo (les exemples présentés dans cette section) et de la chaîne d'analyse présentée ci-dessus nous avons choisi, afin d'obtenir une pureté suffisante, $y_\delta = 1 \times 10^{-5}$, ce qui donne un catalogue de détection de pureté proche de 91%. Ce seuil correspond à 3.9σ de la carte.

Les paragraphes suivants présentent les méthodes utilisées en pratique pour calculer complétude, erreurs photométriques et contamination. Le dernier paragraphe montre comment, à partir des N générations Monte-Carlo, construire le modèle statistique de l'observation. C'est ce modèle qui sera utilisé dans le chapitre 7 pour l'estimation des paramètres des modèles cosmologiques, de formation de structure et de physique tardive des amas. L'exemple présenté dans cette section a été calculé à partir de 100 générations Monte-Carlo de type Olimpo sur 400 deg^2 .

6.2.1 Complétude

Pour calculer la complétude on compare simplement les distributions des amas simulés aux détections "amas" associées à un amas simulé que j'appellerai "amas détectés" dans la suite. (l'algorithme de marquage est présenté au 5.4.2). Ce calcul est possible car l'association entre une détection et un amas simulé assure que :

1. la détection considérée est un amas et donc n'est pas une contamination : $N_{obs}^{cont} = 0$,
2. les erreurs photométriques sont nulles : $P(z_{obs}, Y_{obs} | z_{th}, Y_{th}) = \delta(z_{obs} - z_{th})\delta(Y_{obs} - Y_{th})$ soit : $z_{obs} = z_{th}$ et $Y_{obs} = Y_{th}$ pour l'amas détecté considéré.

On a alors simplement :

$$F(z_{th}, Y_{th}) = \frac{dN_{obs}^{amas}(z_{th}, Y_{th})}{dN_{th}^{amas}(z_{th}, Y_{th})} \quad (6.2)$$

Les figures 6.2.1 et 6.2.2 illustrent le type de résultat obtenu. Les courbes de la

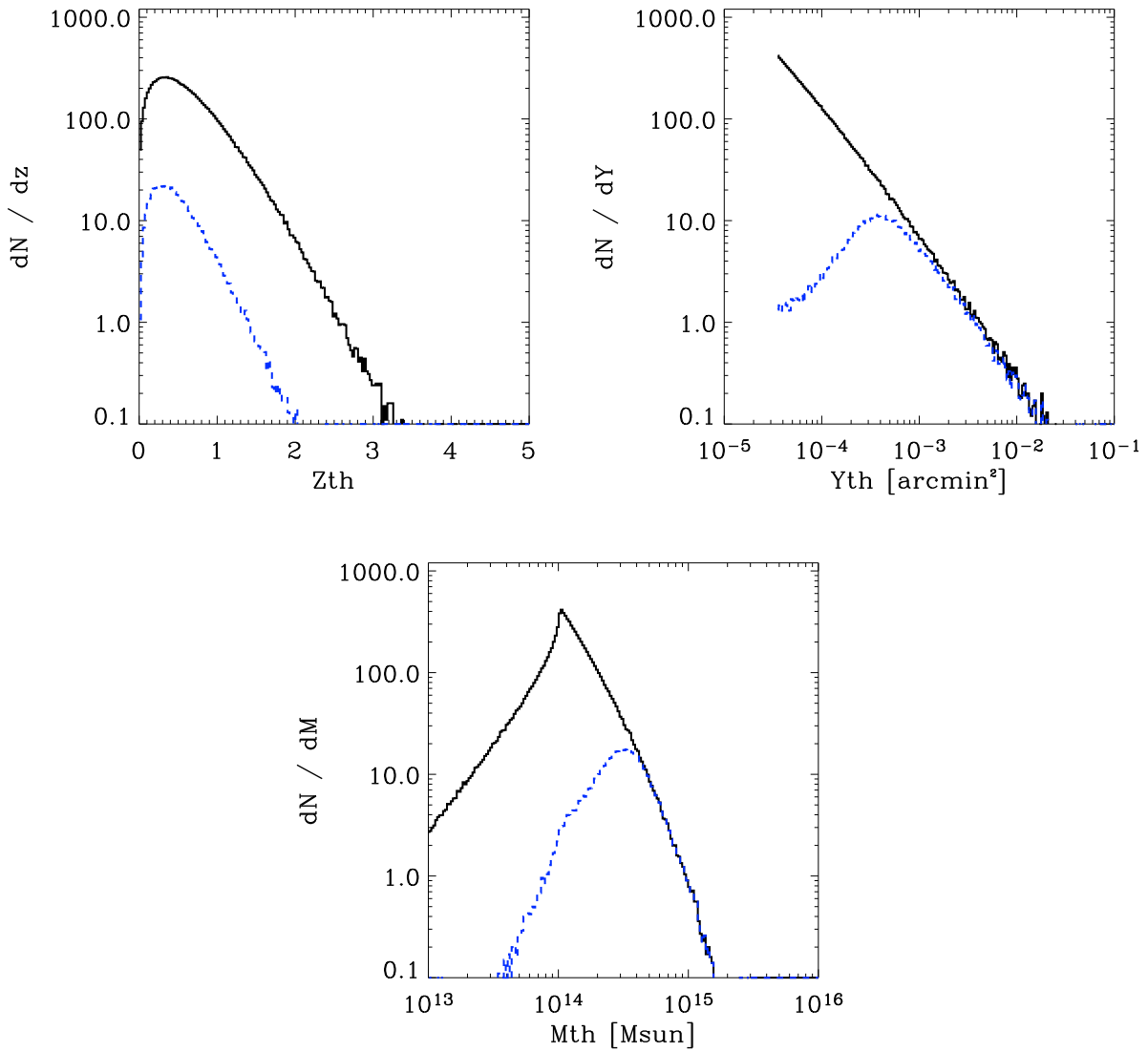


FIG. 6.2.1 – Distribution des amas simulés et détectés en fonction de leur redshift z_{th} , flux Y_{th} et masse M_{th} simulés.

figure 6.2.1 illustrent les distributions simulées et effectivement observées des amas en fonction de leur redshift, de leur flux et de leur masse simulés.

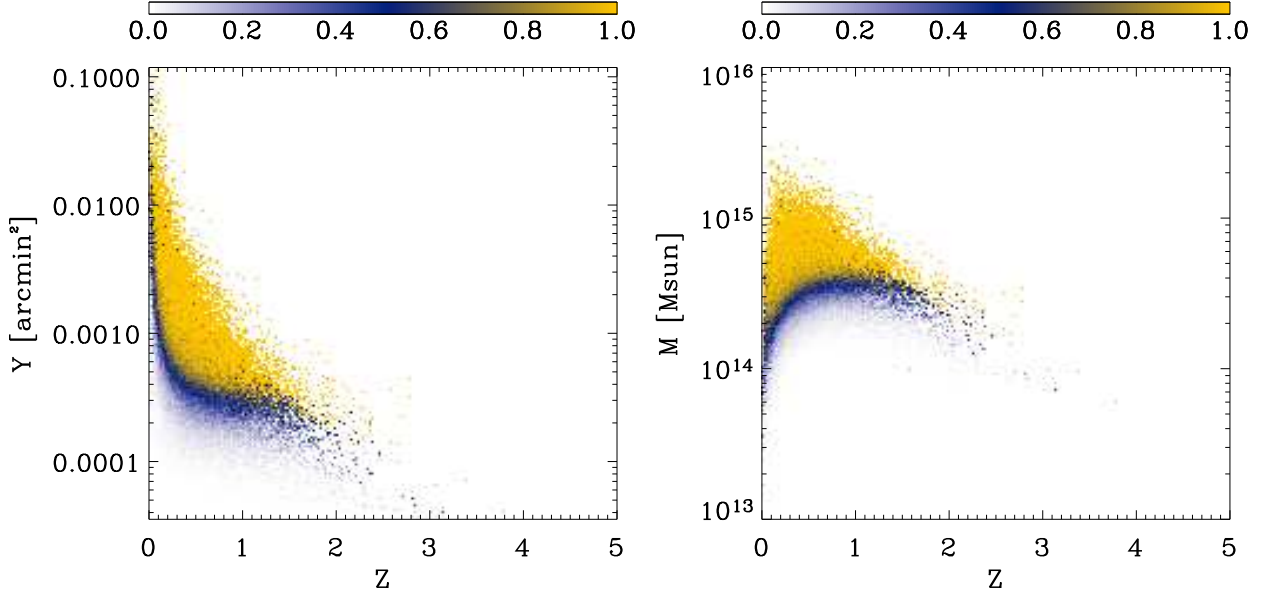


FIG. 6.2.2 – Complétude d’un sondage de type Olimpo en fonction du redshift et du flux simulés des amas (à gauche) et du redshift et de la masse simulée des amas (à droite).

Il est intéressant de remarquer que l’on peut estimer la complétude d’une chaîne de détection en termes d’une caractéristique simulée des amas : leur masse, leur température, leur rayon viriel. Il est évident qu’on ne peut alors pas préciser la contamination ou les erreurs photométriques car ce sont des grandeurs fonction de quantités observables. La complétude donnée en fonction de la masse et du redshift des amas est très intéressante, ces deux quantités étant directement liées à la physique de la formation des grandes structures. La figure 6.2.2 illustre cette complétude.

Modèle statistique. La figure 6.2.2 est produite à partir de générations Monte-Carlo. Nous verrons au chapitre 7 que nous avons besoin, pour calculer les contraintes cosmologiques correctement, de ces mêmes données légèrement modifiées. En effet pour effectuer le calcul :

$$\frac{dN_{obs}^{amas}}{dz_{th}dY_{th}}(z_{th}, Y_{th}) = \frac{dN_{th}^{amas}}{dz_{th}dY_{th}}(z_{th}, Y_{th}) \times F(z_{th}, Y_{th}) \quad (6.3)$$

nous avons besoin de la valeur de F à tous les points (z_{th}, Y_{th}) , or la structure “lacunaire” de la surface de complétude présentée dans la figure 6.2.2 indique que les points, typiquement à haut flux et haute masse n’ont pas été calculés dans le Monte-Carlo car il n’y a simplement pas eu de génération d’amas en ces points. Afin d’estimer la complétude aux points (Y_{th}, z_{th}) non générés par Monte-Carlo

pour les distributions des amas prédites par les modèles de formation de structures, nous introduisons artificiellement des amas dans la région : $1 < z_{th} < 5$ et $3.5 \times 10^{-5} \text{ arcmin}^2 < Y_{th} < 1 \times 10^{-3} \text{ arcmin}^2$. On ajoute à la distribution théorique comptant 9000 amas (pour la cosmologie utilisée pour la génération), 2000 amas (pour chacune des 100 générations) “artificiels”. Ces amas sont introduits de manière à ce que la distribution théorique soit peu perturbée : 1) on vérifie que la confusion ne change pratiquement pas, les amas artificiels ont des tailles apparentes très faibles et influencent peu la capacité de détection des amas “physiques”, 2) la confusion par la distribution théorique domine et permet donc d’avoir une bonne estimation de la complétude pour ces amas artificiels. Les conditions aux points (Y_{th}, z_{th}) générés artificiellement sont identiques à une éventuelle distribution théorique prédisant des amas en ces points. On complète cette génération en constatant que la zone de transition entre une complétude nulle et totale est bien définie pour chaque redshift. Tous les points (Y_{th}, z_{th}) ayant un Y_{th} supérieur à la limite définie par la complétude à 99% et non calculés dans le Monte-Carlo sont fixés à 100%. Un lissage léger permet ensuite de s’affranchir des erreurs statistiques “bin-à-bin” dues à la génération par Monte-Carlo. Ceci est illustré dans la figure 6.2.3.

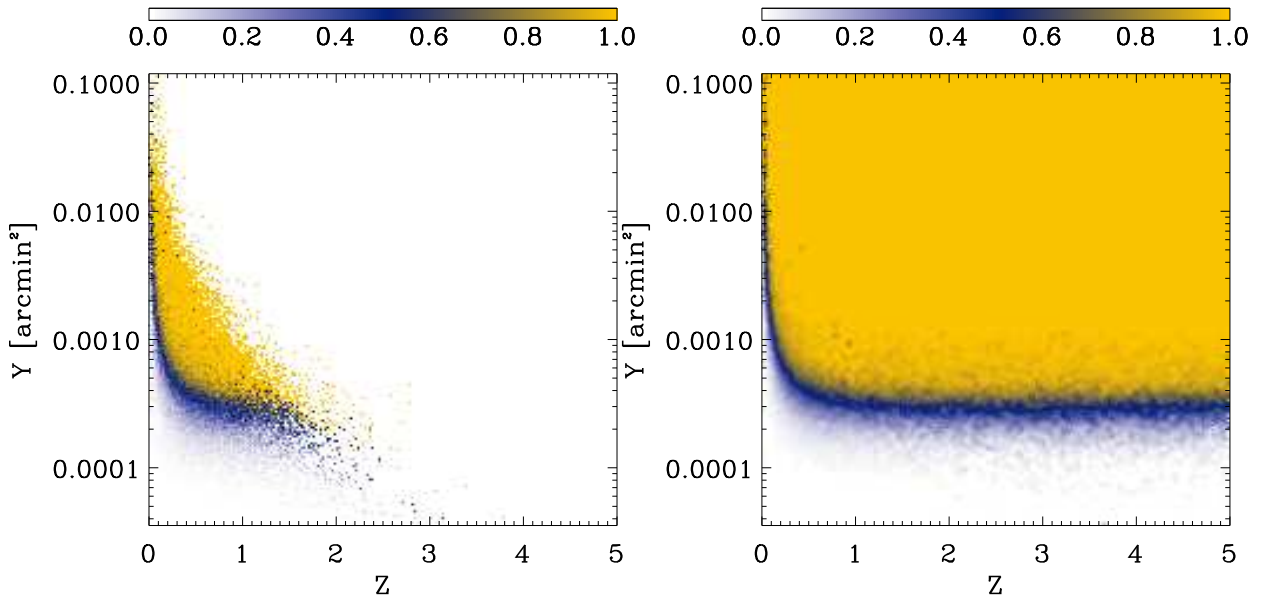


FIG. 6.2.3 – Complétude d’un sondage de type Olimpo tiré du Monte-Carlo (à gauche) et adapté pour le modèle statistique (à droite).

6.2.2 Erreurs photométriques

La photométrie des amas est un travail important et délicat. Souhaitant mettre en place une chaîne complète de simulation-analyse pour une observation SZ grand champ, nous ne nous sommes pas penchés sur l’optimisation d’une photométrie

spécifique aux amas de galaxies. La photométrie du flux des amas est faite grâce au logiciel SExtractor (Bertin and Arnouts, 1996) qui présente l'avantage d'être largement utilisé par la communauté astrophysique, d'être rapide et facilement configurable. J'ai commencé à travailler sur un algorithme de photométrie basé sur la connaissance, *a priori*, du profil de luminosité des amas : ce travail est encore trop préliminaire pour être présenté ici. La photométrie SExtractor utilisée est : FLUX_AUTO. C'est une photométrie basée sur un algorithme assez général qui s'adapte à la taille de la détection, ce qui permet, *a priori*, d'estimer correctement leur flux intégré.

On cherche à estimer les caractéristiques observées à partir des caractéristiques simulées des amas, c'est à dire :

$$\frac{dP}{dY_{obs}}(Y_{obs}|z_{th}, Y_{th}) \quad (6.4)$$

et

$$\frac{dP}{dz_{obs}}(z_{obs}|z_{th}, Y_{th}) \quad (6.5)$$

Le calcul de ces deux quantités caractérise complètement les erreurs photométriques de la fonction de sélection. Dans les paragraphes suivants nous ferons certaines hypothèses simplificatrices. Un fois ces quantités estimées, le théorème de Bayes donne :

$$\frac{dN_{obs}^{amas}}{dz_{obs}dY_{obs}}(z_{obs}, Y_{obs}) = \iint \frac{dN_{obs}^{amas}}{dz_{th}dY_{th}}(z_{th}, Y_{th}) \cdot \frac{dP}{dY_{obs}}(Y_{obs}|Y_{th}) \times \frac{dP}{dz_{obs}}(z_{obs}|z_{th}) dY_{th}dz_{th} \quad (6.6)$$

Nous avons choisi de ne pas considérer la taille apparente des amas comme observable : la figure 6.2.4 illustre la difficulté pour estimer cette grandeur des sources brillantes avec l'algorithme de photométrie utilisé. Les erreurs sur le calcul de la taille sont trop importantes pour pouvoir espérer en tirer beaucoup d'information sur la physique des amas.

6.2.2.1 Estimation de Y_{obs}

Afin de construire le modèle statistique de la chaîne de détection on s'intéresse au rapport : $\ln Y_{obs} / \ln Y_{th}$. Les figures 6.2.5 et 6.2.6 montrent un exemple de distribution. Ces figures suggèrent plusieurs remarques :

1. le rapport $\ln Y_{obs} / \ln Y_{th}$ est quasiment indépendant de z_{th} . Nous supposons dans la suite que le flux reconstruit est indépendant du redshift de l'amas détecté : Y_{obs} est fonction uniquement de Y_{th} . Cependant on constate à bas redshift (< 0.2) un léger biais positif. Nous n'en tenons pas compte pour le moment.
2. Il existe une erreur systématique importante sur le flux estimé à bas flux.
3. Les distributions $\ln Y_{obs}(Y_{th})$ peuvent être modélisées correctement par des distributions gaussiennes.

On peut supposer deux origines aux erreurs photométriques : le bruit de mesure et le bruit de confusion, c'est à dire les effets de superposition des amas.

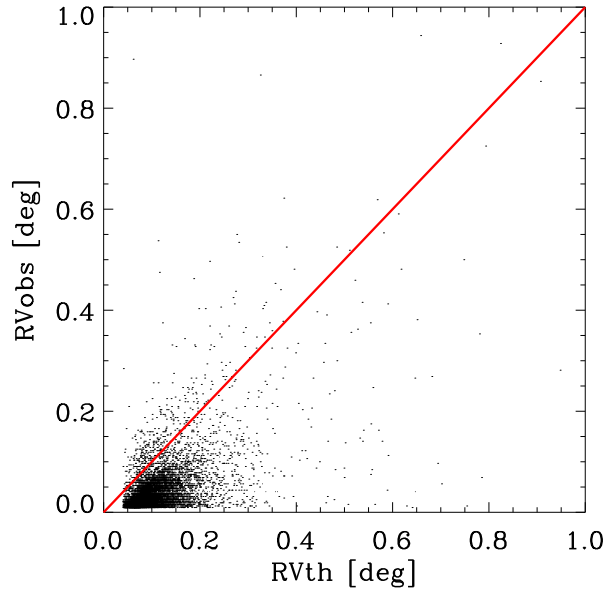


FIG. 6.2.4 – Reconstruction du rayon viriel apparent des amas détectés par SExtractor. L’abscisse donne le rayon viriel simulé apparent (RVth) en degrés et l’ordonnée le rayon viriel apparent (RVobs) reconstruit en degrés.

Le biais observé dans le flux estimé Y_{obs} est dû à l’algorithme de photométrie, car nous nous sommes placés dans un régime où les erreurs de marquage sont négligeables. Nous en tenons compte dans le modèle statistique de l’instrument.

Modèle statistique. En supposant les distributions de $\ln Y_{obs}(Y_{th})$ gaussienne, le modèle statistique de la chaîne de détection, en terme d’erreur photométrique sur Y_{obs} , est donné par les grandeurs : $\ln \bar{Y}_{obs}(Y_{th})$ et $\sigma_{\ln Y_{obs}}(Y_{th})$. Par génération Monte-Carlo on obtient facilement une estimation de ces fonctions donnée dans la figure 6.2.7. La connaissance de Y_{obs} est modélisée par la distribution suivante :

$$\frac{dP}{d \ln Y_{obs}}(\ln Y_{obs}|Y_{th}) = \frac{1}{\sqrt{2\pi} \sigma_{\ln Y_{obs}}(Y_{th})} \exp \left[-\frac{(\ln Y_{obs} - \ln \bar{Y}_{obs}(Y_{th}))^2}{2 \sigma_{\ln Y_{obs}}^2(Y_{th})} \right] \quad (6.7)$$

Cette distribution permet d’estimer la valeur de Y_{obs} connaissant Y_{th} . La figure 6.2.7 résume le modèle statistique des erreurs photométriques.

La mise en place d’algorithmes de photométrie adaptés aux amas permettrait d’utiliser l’information sur la taille des amas détectés. On pourrait envisager, en faisant l’hypothèse que le gaz des amas suit, “en moyenne”, un profil beta, d’utiliser un algorithme de photométrie basée sur un ajustement de profil beta au profil des détections. Ceci permettrait de pouvoir estimer le rayon de coeur et le flux intégré des amas de manière plus précise que l’algorithme utilisé ici. Un algorithme utilisant le principe des *matched filter* permettrait aussi d’estimer le rayon de coeur des amas

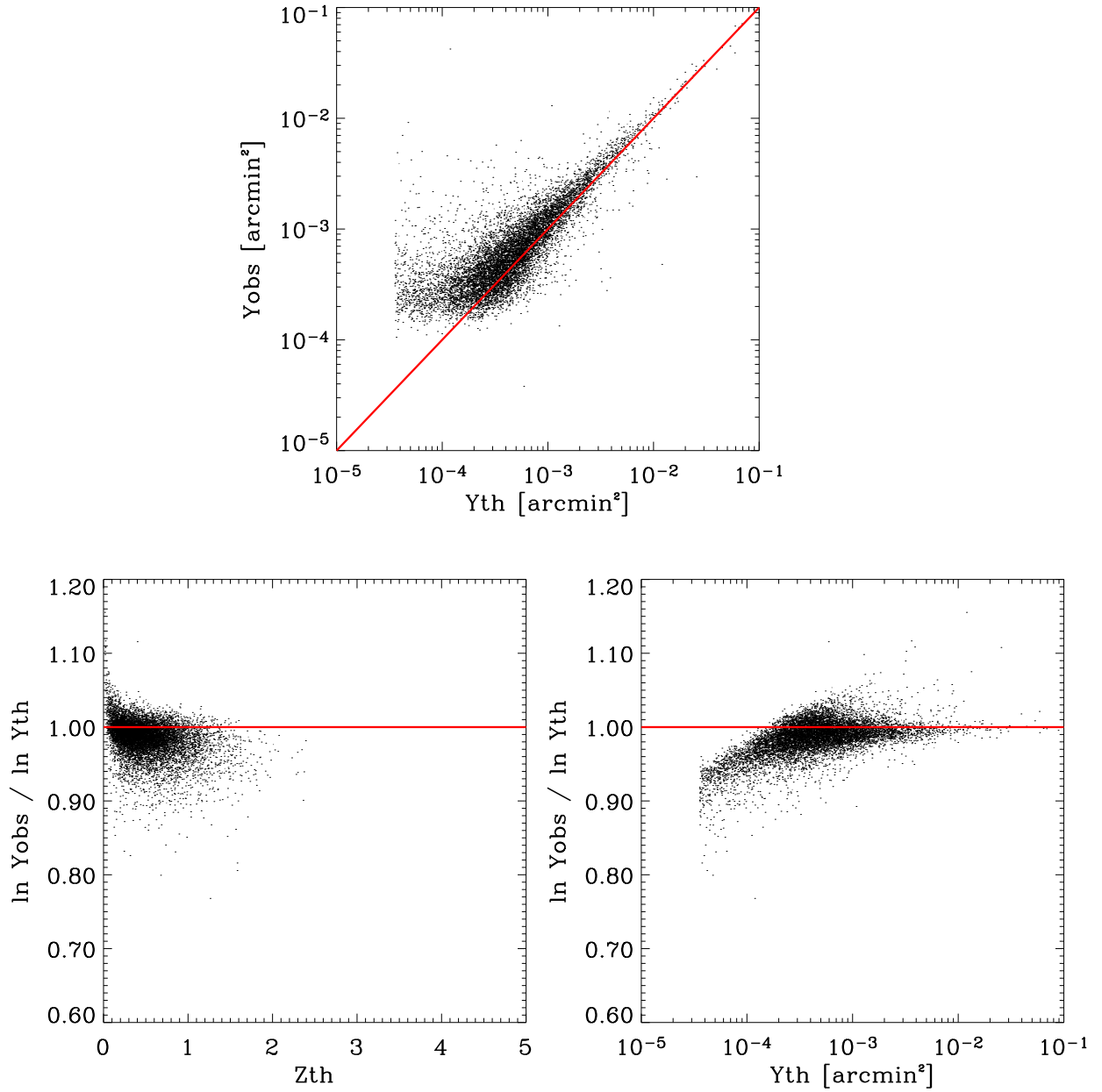


FIG. 6.2.5 – Estimation du paramètre de Compton intégré Y_{obs} pour les amas détectés. La figure en haut donne le paramètre Y_{obs} en fonction du paramètre Y_{th} . Les deux figures du bas donnent le rapport $\ln Y_{obs} / \ln Y_{th}$ en fonction de z_{th} (à gauche) et Y_{th} à droite. Les lignes rouges indiquent l'ensemble des points où $Y_{obs} = Y_{th}$. Dans ces figures seules 20 des 100 générations du Monte-Carlo sont présentées.

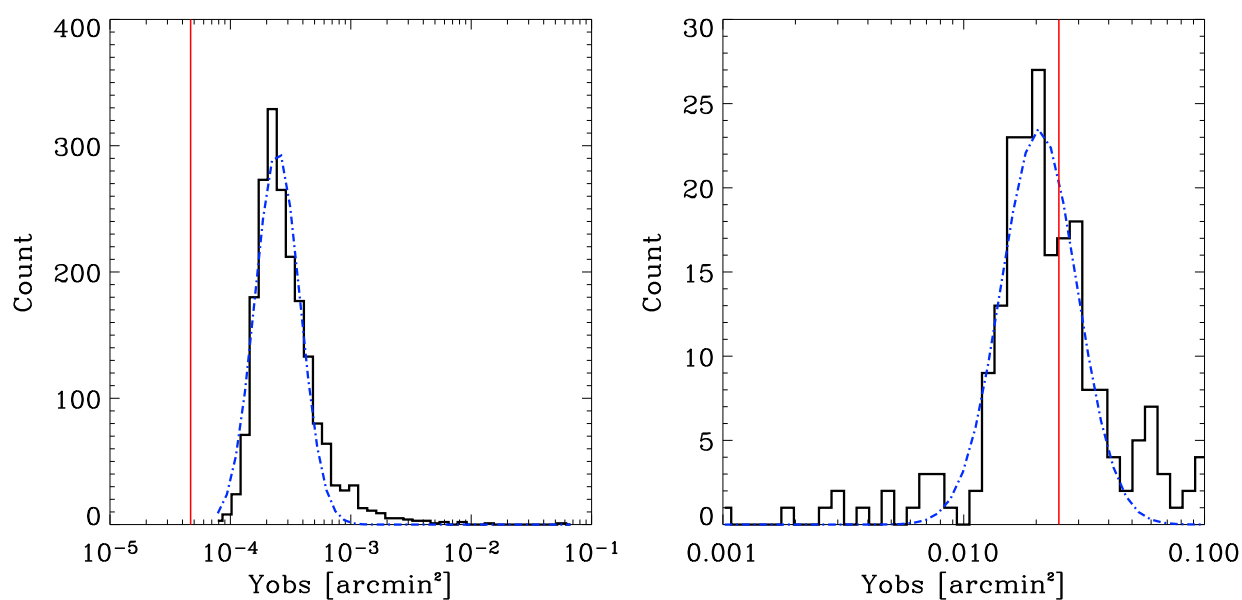


FIG. 6.2.6 – Distributions, pour deux flux théoriques donnés ($Y_{th} = 5 \times 10^{-5} \text{ arcmin}^2$ à gauche et $Y_{th} = 2.5 \times 10^{-2} \text{ arcmin}^2$ à droite), du flux observé Y_{obs} . La courbe discontinue bleue indique le modèle gaussien utilisé dans le modèle statistique et le trait continu rouge donne la valeur du flux théorique Y_{th} considéré.

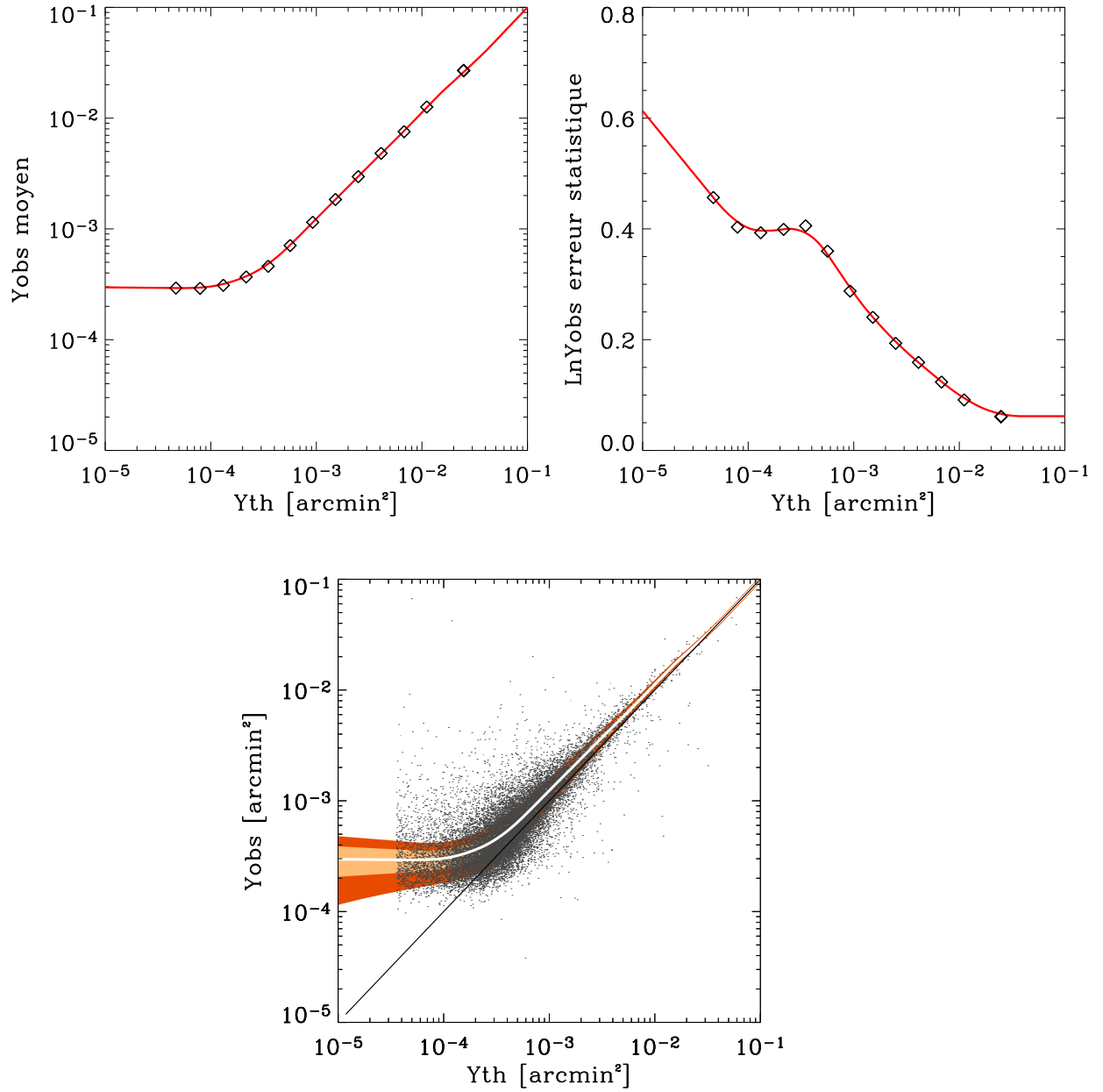


FIG. 6.2.7 – En haut à gauche : \bar{Y}_{obs} en fonction de Y_{th} . En haut à droite : $\sigma_{\ln Y_{obs}}$ en fonction de Y_{th} . En bas : figure résumant le modèle statistique des erreurs photométriques sur l'estimation de Y_{obs} . En trait continu blanc \bar{Y}_{obs} tiré du modèle statistique, les zones oranges correspondent à $\sigma_{\ln Y_{obs}}$ (orange clair) et $2\sigma_{\ln Y_{obs}}$ (orange foncé) tiré du modèle statistique. Le nuage de points présente les données des générations (20 sur les 100) du Monte-Carlo.

et leur flux intégré (Melin, 2004). Cependant la donnée du rayon de coeur ne permet de retrouver le rayon viriel des amas qu'à une relation d'échelle près, l'information relative aux modèles cosmologiques reste donc très faible.

6.2.2.2 Estimation de z_{obs}

Nous avons considéré que le redshift des amas constitue une observable de notre modèle d'expérience. Bien entendu cette grandeur ne pourra être mesurée à partir des sondages SZ seuls (bien que des travaux montrent qu'on pourrait estimer le redshift des amas par des observations conjointes X, SZ). Cependant nous supposons pouvoir disposer ultérieurement des redshifts d'une fraction des détections dans le champ observé. Nous étudions dans le chapitre 7 l'impact d'un tel suivi en termes de contraintes sur les modèles cosmologiques.

Afin de construire le modèle statistique des erreurs de mesure sur le redshift photométrique nous avons besoin des informations suivantes : $\bar{z}_{obs}(z_{th}, Y_{th})$ et $\sigma_{z_{obs}}(z_{th}, Y_{th})$, c'est à dire la valeur moyenne du redshift observé connaissant le redshift réel et l'erreur standard sur cette mesure (erreur systématique et erreur statistique).

Nous faisons les hypothèses simplificatrices suivantes : $\bar{z}_{obs} = z_{th}$ et $\sigma_{z_{obs}} \sim \text{constant}$. C'est à dire que l'on considère que le redshift est mesuré indépendamment du flux de la détection, que sa valeur moyenne est la valeur réelle (pas de biais d'estimation) et que l'erreur standard est constante : le redshift photométrique dépend surtout du nombre de galaxies utilisées pour l'estimer. Nous supposons pouvoir mesurer le redshift d'une dizaine de galaxies pour chaque amas, soit une erreur : $\sigma_{z_{obs}} = 0.1$ quel que soit le redshift.

Modèle statistique. En pratique nous tabulons la courbe : $\sigma_{z_{obs}}(z_{th})$ une fois pour toute. Pour des raisons algorithmiques, nous avons modélisé cette courbe en prenant pour $z_{th} < 0.1$, $\sigma_{z_{obs}} \propto z_{obs}$.

On a ainsi :

$$\frac{dP}{dz_{obs}}(z_{obs}|z_{th}) = \frac{1}{\sqrt{2\pi} \sigma_{z_{obs}}} \exp \left[-\frac{(z_{obs} - \bar{z}_{obs}(z_{th}))^2}{2 \sigma_{z_{obs}}^2} \right] \quad (6.8)$$

avec $\bar{z}_{obs} = z_{th}$ et $\sigma_{z_{obs}} = 0.1$.

6.2.3 Les contaminations

On introduit les contaminations en ajoutant à la distribution théorique des amas, multipliée par la complétude, "convoluée" par les erreurs photométriques, la distribution des contaminations (voir l'équation 6.1). La prise en compte des contaminations dépend du suivi en redshift considéré.

En supposant pouvoir mesurer le redshift photométrique de 100% des détections dans le champ observé, il sera possible d'identifier et donc d'éliminer les contaminations de nos mesures. En effet un suivi optique en parallèle pourra nous révéler si une détection est due à un amas (présence de galaxies dans la direction de la détection) ou à du bruit instrumental (pas de galaxies dans la direction de la détection). Ainsi

avec 100% des redshifts mesurés on peut espérer obtenir un catalogue sans contamination et la fonction de sélection fait alors intervenir uniquement la complétude et les erreurs photométriques.

Si on suppose maintenant pouvoir mesurer le redshift d'une fraction seulement du catalogue de détection, alors on classe les détections en trois catégories : les amas détectés (détection ayant été examinée par le programme de suivi et ayant révélé la présence d'un amas), les contaminations (source ayant été observée lors du suivi mais n'ayant pas révélé d'amas), et les autres détections (dont on ne sait pas si ce sont des amas ou des contaminations). Cette dernière catégorie est prise en compte et on doit donc connaître la distribution en flux observé Y_{obs} des contaminations :

$$\frac{dN_{obs}^{cont}}{dY_{obs}}(Y_{obs})$$

On peut déterminer cette fonction à partir de la simulation Monte-Carlo. Le résultat est illustrée dans la figure 6.2.8.

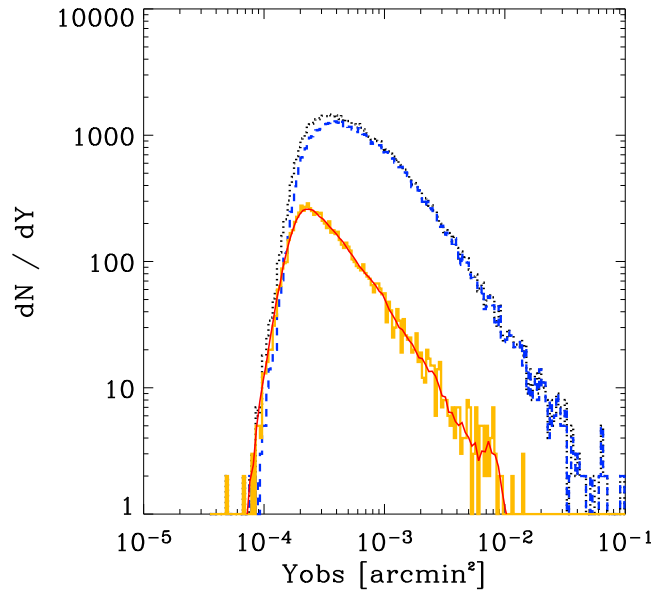


FIG. 6.2.8 – Distribution des contaminations en fonction de leur flux estimé Y_{obs} . La courbe rouge continue donne le modèle statistique de la distribution des contaminations dN_{obs}^{cont}/dY_{obs} tandis que l'histogramme en orange donne la distribution Monte-Carlo à partir de laquelle il a été construit. La courbe bleue en trait discontinu donne la distribution des amas détectés dN_{obs}^{amas}/dY_{obs} . La courbe noire en pointillés donne la distribution des détections dN_{obs}^{det}/dY_{obs} . Ces courbes intégrées donnent une pureté de 91%.

6.2.4 Nombre effectif de détections

La complétude du sondage permet de calculer le nombre moyen d'amas détectés : \bar{N}_{obs}^{amas} . La distribution des contaminations permet de calculer le nombre moyen de contaminations détectées : \bar{N}_{obs}^{cont} . Le nombre moyen de détections est alors simplement : $\bar{N}_{obs}^{det} = \bar{N}_{obs}^{amas} + \bar{N}_{obs}^{cont}$. Qu'en est-il du nombre de détections (amas ou contamination) effectif pour une observation donnée ?

Les distributions des nombres "effectifs" N de contaminations et d'amas ne sont pas poissonniennes, bien que le nombre effectif d'amas simulé dans une génération soit issu d'un tirage poissonien autour de la valeur moyenne théorique prédite par le modèle de formation de structure. Ces distributions sont extrêmement importantes pour calculer les contraintes sur les paramètres cosmologiques présentées au chapitre 7. La figure 6.2.9 illustre les distributions obtenues par Monte-Carlo. Nous modélisons les distributions des nombres effectifs de détections par des distributions gaussiennes (courbes rouges-grises de la figure) :

– pour les amas détectés :

$$\frac{dP}{dN_{obs}^{amas}} = \frac{1}{\sqrt{2\pi}\sigma_{N_{obs}^{amas}}} \exp \left[-\frac{1}{2} \left(\frac{N_{obs}^{amas} - \bar{N}_{obs}^{amas}}{\sigma_{N_{obs}^{amas}}} \right)^2 \right] \quad (6.9)$$

– pour les contaminations :

$$\frac{dP}{dN_{obs}^{cont}} = \frac{1}{\sqrt{2\pi}\sigma_{N_{obs}^{cont}}} \exp \left[-\frac{1}{2} \left(\frac{N_{obs}^{cont} - \bar{N}_{obs}^{cont}}{\sigma_{N_{obs}^{cont}}} \right)^2 \right] \quad (6.10)$$

– pour toutes les détections :

$$\frac{dP}{dN_{obs}^{det}} = \frac{1}{\sqrt{2\pi}\sigma_{N_{obs}^{det}}} \exp \left[-\frac{1}{2} \left(\frac{N_{obs}^{det} - \bar{N}_{obs}^{det}}{\sigma_{N_{obs}^{det}}} \right)^2 \right] \quad (6.11)$$

On décrit ces distributions à partir d'une généralisation d'une distribution de poisson en posant :

$$\sigma_{N_{obs}^{amas}} = A^{amas} \sqrt{\bar{N}_{obs}^{amas}} \quad (6.12)$$

$$\sigma_{N_{obs}^{cont}} = A^{cont} \sqrt{\bar{N}_{obs}^{cont}} \quad (6.13)$$

$$\sigma_{N_{obs}^{det}} = A^{det} \sqrt{\bar{N}_{obs}^{det}} \quad (6.14)$$

En supposant le modèle statistique de la chaîne de détection indépendant du modèle cosmologique utilisé pour le calculer, on suppose les paramètres A constant et on a $A^{amas} = 3.03$, $A^{cont} = 3.04$ et $A^{det} = 3.70$. On peut imaginer construire dans un travail futur une modélisation plus fine de ces distributions, notamment en introduisant une dépendance en \bar{N} dans les paramètres A .

On voit sur la figure 6.2.9 que les distributions présentent une queue qui s'étire vers les grands nombres de détections. Une étude détaillée a montré une corrélation

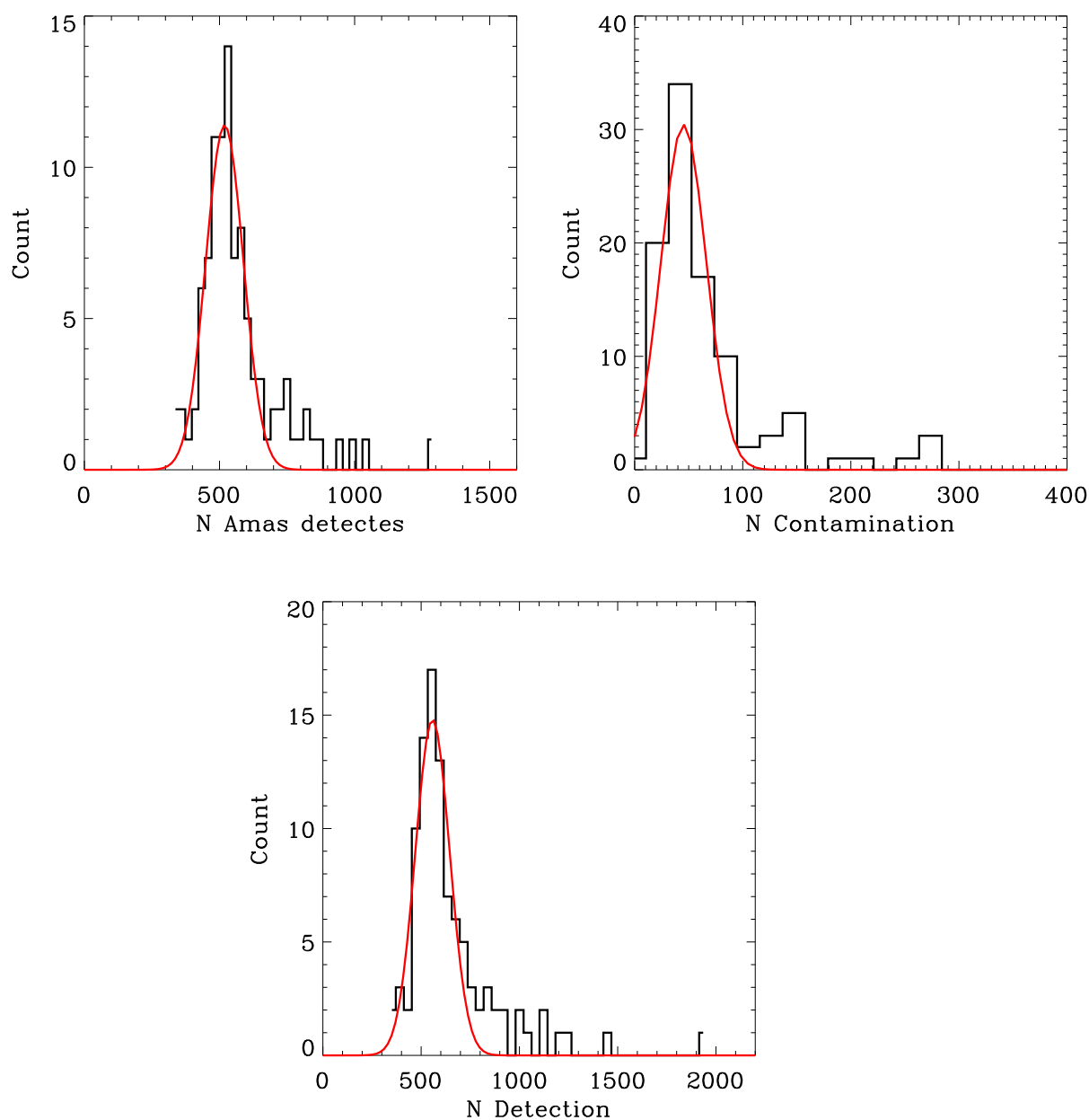


FIG. 6.2.9 – Distributions du nombre d’amas détectés (en haut à gauche), du nombre de contaminations (en haut à droite) et du nombre de sources détectées (en bas) pour 100 générations complètes simulation-détection de type “Olimpo”. Les courbes rouges sont les meilleurs ajustement de distributions gaussiennes.

entre un nombre “anormal” (en terme de distribution gaussienne) de contamination et un nombre “anormal” d’amas détectés. Ces nombres “anormaux” de détections sont dus à des erreurs dans la renormalisation des cartes SZ reconstruites. En effet, dans certains cas, on constate que le flux des plus gros amas SZ est sous estimé conduisant à une renormalisation de la carte erronée. Le seuil de détection étant fixe et indépendant de la carte considérée, ceci conduit à une surestimation du nombre de détections par carte. Le choix de modéliser ces distributions par des gaussiennes reste donc valable, les cas anormaux étant compris. Un travail plus poussé de photométrie ou de renormalisation des cartes reconstruites permettra de s’en affranchir (par exemple, nous avons utilisé une régression linéaire par χ^2 qui est un estimateur connu pour être sensible aux valeurs anormales des distributions, on peut imaginer construire une régression linéaire plus adaptée) .

6.3 Validité du modèle statistique

6.3.1 Complétude, photométrie et contaminations

Nous avons présenté dans ce chapitre tous les ingrédients du modèle statistique que nous avons mis en place pour Olimpo permettant, à partir de la distribution théorique des amas, de calculer la distribution effectivement observée des détections. Ce modèle n'est pas spécifique à Olimpo et convient pour d'autres expériences bolométriques multi-fréquences. Le modèle statistique pour Olimpo, présenté ici, comprend :

1. la fonction de complétude $F(z_{th}, Y_{th})$ qui permet de calculer, la distribution des amas détectés en fonction de leur flux et de leur redshift théorique (Y_{th} et z_{th}) :

$$\frac{dN_{obs}^{amas}}{dz_{th}dY_{th}}(z_{th}, Y_{th}) = \frac{dN_{th}^{amas}}{dz_{th}dY_{th}}(z_{th}, Y_{th}) \times F(z_{th}, Y_{th}) \quad (6.15)$$

2. les erreurs photométriques $P(z_{obs}, Y_{obs}|z_{th}, Y_{th})$ qui permettent de calculer la distribution des amas détectés en fonction de leur flux estimé Y_{obs} et redshift estimé z_{obs} . Le théorème de Bayes permet en effet de dire que

$$\frac{dN_{obs}^{amas}}{dz_{obs}dY_{obs}}(z_{obs}, Y_{obs}) = \iint \frac{dN_{obs}^{amas}}{dz_{th}dY_{th}}(z_{th}, Y_{th}) \cdot \frac{dP}{dY_{obs}}(Y_{obs}|Y_{th}) \times \frac{dP}{dz_{obs}}(z_{obs}|z_{th}) dY_{th}dz_{th} \quad (6.16)$$

3. la distribution des contaminations $\frac{dN_{obs}^{cont}}{dY_{obs}}(Y_{obs})$ qui permet de calculer la distribution des détections, Si on mesure les redshift (pas de contaminations), on a

$$\frac{dN_{obs}^{det}}{dz_{obs}dY_{obs}}(z_{obs}, Y_{obs}) = \frac{dN_{obs}^{amas}}{dz_{obs}dY_{obs}}(z_{obs}, Y_{obs}) \quad (6.17)$$

si on ne mesure pas les redshift :

$$\frac{dN_{obs}^{det}}{dY_{obs}}(Y_{obs}) = \frac{dN_{obs}^{amas}}{dY_{obs}}(Y_{obs}) + \frac{dN_{obs}^{cont}}{dY_{obs}}(Y_{obs}) \quad (6.18)$$

Il est possible de vérifier la validité du modèle statistique construit à partir du Monte-Carlo en comparant, pour le même modèle cosmologique \vec{P} , les distributions issues du Monte-Carlo, des amas détectés, et des contaminations, aux distributions obtenues par l'application du modèle statistique. Cette comparaison est illustrée dans les figures 6.3.1 et 6.3.2. La figure 6.3.1 prouve le bon fonctionnement du modèle statistique, en termes de complétude (deuxième ligne) et d'erreur photométrique (troisième ligne). On remarque sur la troisième ligne, colonne de droite, que le profil est plus élargi en redshift (abscisse) que pour le Monte-Carlo. Ceci est dû au fait que le modèle statistique d'erreurs photométriques en redshift n'est pas déduit du Monte-Carlo dans lequel on le considère parfaitement connu mais il est introduit par ailleurs. La distribution du Monte-Carlo ne contient pas d'erreur photométrique en redshift. La distribution obtenue grâce au modèle statistique en contient, ce qui correspond à une convolution par un noyau gaussien et se manifeste par une "dilution" de la distribution. Il y a deux autres différences notables :

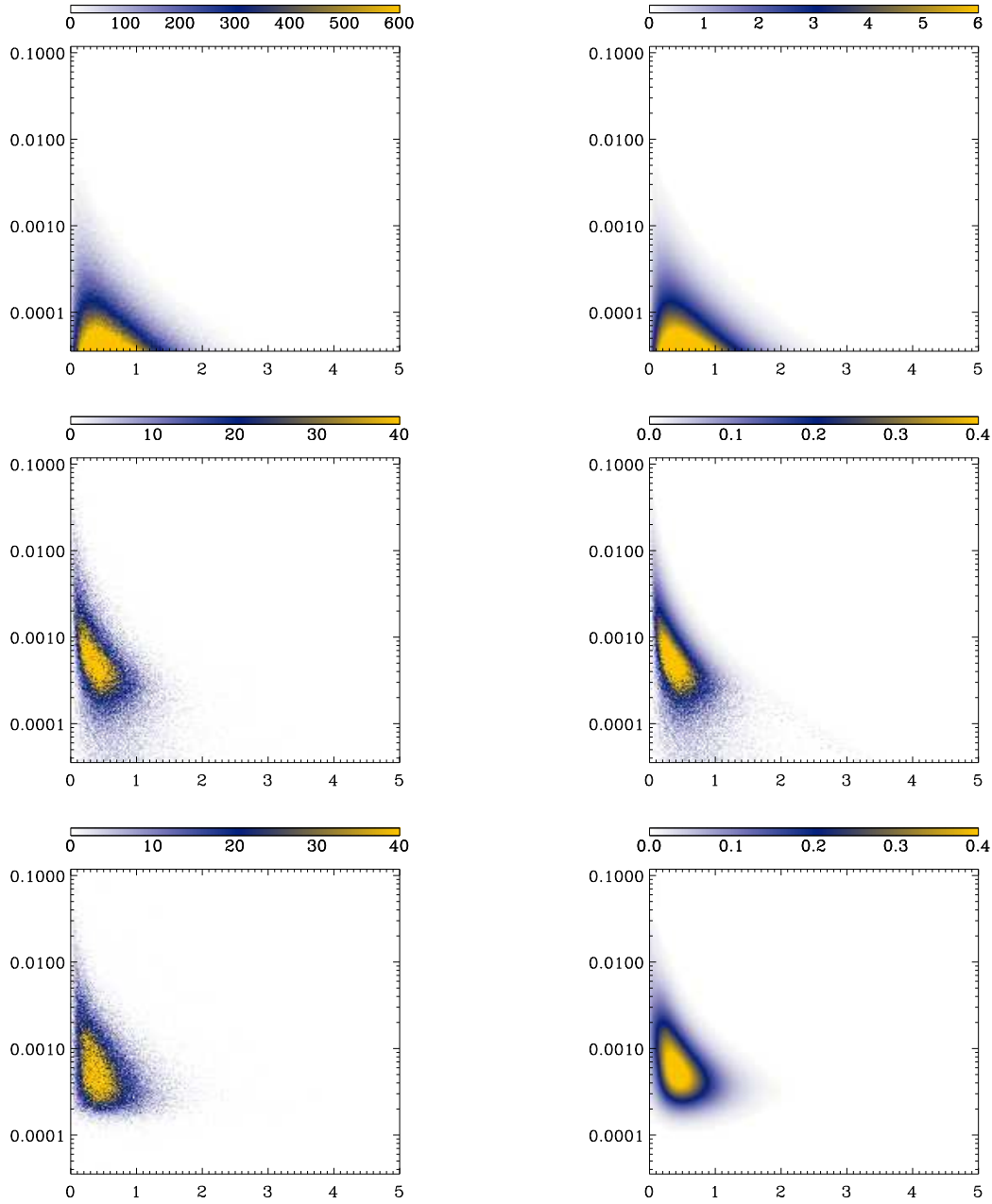


FIG. 6.3.1 – Distribution des amas détectés en fonction du redshift (abscisse) et du paramètre de Compton intégré (ordonnée) pour le même modèle cosmologique. La colonne de gauche présente les distributions issues des 100 générations du Monte-Carlo tandis que la colonne de droite donne les distributions obtenues par application du modèle statistique présenté dans ce chapitre. La première ligne présente les modèles théoriques simulés : $\frac{dN_{th}^{amas}}{dz_{th}dY_{th}}(z_{th}, Y_{th})$. La deuxième ligne présente les distributions observées d'amas en fonction de leur flux et redshift simulé : $\frac{dN_{obs}^{amas}}{dz_{th}dY_{th}}(z_{th}, Y_{th})$. La troisième ligne présente les distributions observées des amas en fonction de leur flux et redshift observé : $\frac{dN_{obs}^{amas}}{dz_{obs}dY_{obs}}(z_{obs}, Y_{obs})$.

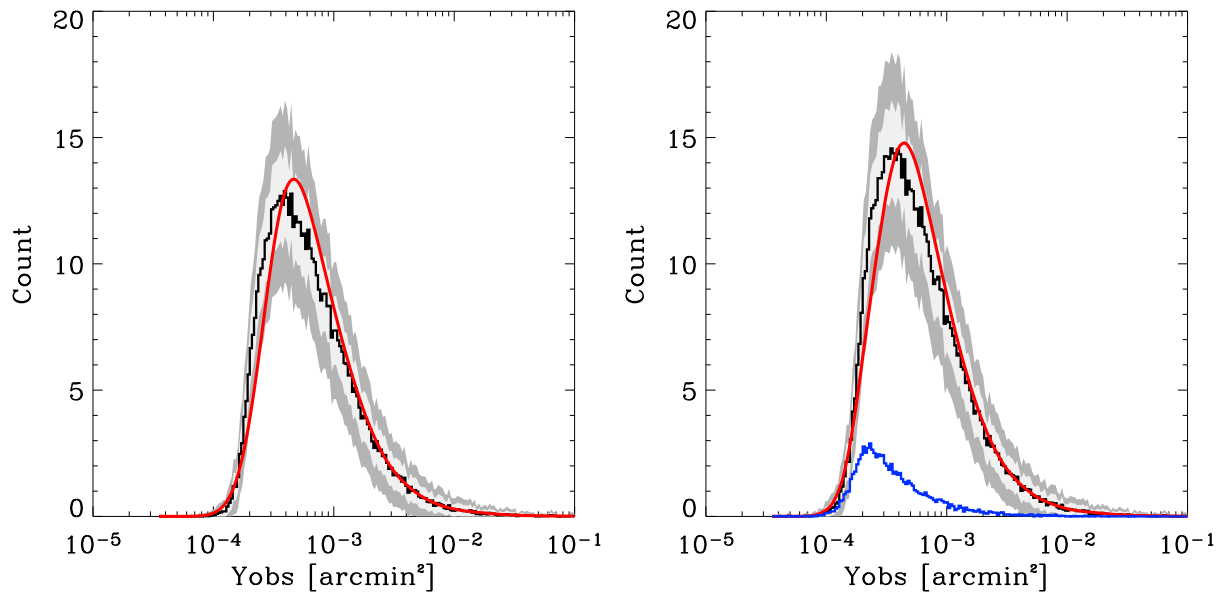


FIG. 6.3.2 – A gauche : distribution en fonction de Y_{obs} des amas détectés. L'histogramme en trait épais donne la distribution extraite des 100 générations Monte-Carlo, le trait plein, rouge et fin donne la distribution obtenue par l'utilisation du modèle statistique. A droite : distribution en fonction de Y_{obs} des détections. L'histogramme en trait épais donne la distribution extraite des 100 générations du Monte-Carlo, le trait continu, rouge donne la distribution obtenue par l'utilisation du modèle statistique. Dans les deux figures les zones grisées donnent, à un (gris clair) et deux sigmas (gris foncé), les erreurs statistiques du Monte-Carlo pour *une seule* génération.

- à bas redshift les courbes d'iso-densité pour le modèle statistique semblent s'écarter plus rapidement de la droite $z = 0$ lorsque le flux diminue. Ceci est dû au fait que l'on ne modélise pas le biais photométrique en fonction du redshift, pourtant non-nul pour les bas redshift (et uniquement pour ceux-ci). Ce biais est visible sur la figure 6.2.5 en bas à gauche.
- Le deuxième point où le modèle semble imprécis concerne le gradient de la transition entre zéro et 100% de densité, à un redshift donné, lorsque le flux augmente. On constate en effet que notre modèle statistique présente une montée un peu plus douce que les distributions Monte-Carlo. Ceci s'explique par la modélisation des erreurs photométriques par des gaussiennes alors que les distributions sont plus pentues à bas flux.

La figure 6.3.2 (gauche) donne la même information que la troisième ligne de la figure 6.3.1 en fonction uniquement du flux observé. L'histogramme en noir donne la distribution des amas détectés pour les 100 générations du Monte-Carlo, le trait continu, rouge donne la distribution obtenue grâce au modèle statistique. On constate une petite erreur systématique entre les deux distributions. Les zones grisées donnent les erreurs statistiques à une (gris clair) et deux (gris foncé) déviations standards (statistique poissonnienne) pour *une seule* génération, c'est à dire pour une observation typique d'Olimpo. On constate que l'erreur systématique est inférieure à l'erreur statistique à un sigma pour tous les flux supérieurs à 3×10^{-4} Sr et à deux sigmas pour tous les flux considérés. Ceci nous confirme le fonctionnement correct du modèle statistique. Pour aller plus loin, on pourrait utiliser des distributions légèrement non-gaussiennes avec des moments d'ordre supérieurs à deux non nuls pour ajuster les distributions Monte-Carlo. La figure de droite présente les mêmes données mais pour la distribution des détections (amas et contamination). On constate que le modèle statistique des contaminations est lui aussi valide : il n'y a pas d'erreur systématique introduite par la prise en compte des contaminations dans le modèle statistique.

6.3.2 Modèle statistique et confusion

Les différentes quantités caractérisant le modèle statistique de la chaîne de détection (complétude, contaminations, erreurs photométriques et nombre effectif de détections) sont supposées indépendantes du cadre cosmologique ayant servi à les calculer. Ceci n'est qu'une approximation. Il y a en réalité une certaine dépendance dans la mesure où la confusion des sources peut dégrader les performances de la chaîne de détection pour des densités de surface d'amas importantes.

Le bruit de confusion est dû aux sources SZ non détectées. On peut estimer son amplitude pour chaque source détectée par la relation :

$$\sigma_c^2(\theta_{det}) = \theta_{det} \int Y_{th}^2 \frac{dN_{th}^{nondet}}{dY_{th}d\Omega} dY_{th} \quad (6.19)$$

où θ_{det} est la taille apparente de la détection. La distribution $\frac{dN_{th}^{nondet}}{dY_{th}d\Omega}$ des amas simulés non-détectés par unité de surface du ciel observé est parfaitement connue :

$$\frac{dN_{th}^{nondet}}{dY_{th}} = \frac{dN_{th}}{dY_{th}} - \frac{dN_{th}^{det}}{dY_{th}} \quad (6.20)$$

Cette relation provient du calcul du bruit de confusion pour la photométrie des sources infrarouges de FIRBACK décrit dans la thèse de H. Dole (Dole, 2000). Pour les sources de FIRBACK le calcul est plus simple car la taille des sources est constante : c'est la taille du lobe instrumental. Dans notre cas les amas sont des sources dont la taille apparente excède parfois la taille du lobe instrumental. L'estimation du bruit de confusion par cette méthode et par Monte-Carlo est un travail qui devra être abordé par la suite.

En pratique le niveau de confusion est contrôlé dans le Monte-Carlo en introduisant, pour une même cosmologie, c'est à dire pour une même distribution théorique d'amas en redshift et flux, un nombre total d'amas variable. Dans les exemples donnés ici, la référence est la génération avec le nombre moyen d'amas prévu par le modèle de formation de structure (~ 9000 , modèle statistique *MS_1.0*), nous avons ensuite construit un modèle statistique en introduisant 1.5 fois plus d'amas (~ 13500 , modèle *MS_1.5*) et 0.75 fois moins (~ 6700 , modèle *MS_0.75*). Les résultats présentés ici illustrent donc la dépendance du modèle statistique vis-à-vis du nombre d'amas moyen prévu dans un cadre cosmologique donné. La dépendance vis-à-vis de leur distribution en flux et redshift n'est pas illustrée. Un nombre de 9000 amas générés correspond à une valeur de σ_8 de 0.85, 13500 à $\sigma_8 \sim 0.95$ et 6700 à $\sigma_8 \sim 0.75$ (tous les autres paramètres étant égaux par ailleurs).

Complétude. La figure 6.3.3 illustre pour différents niveaux de confusion la surface de complétude calculée par Monte-Carlo. Les contours donnent la complétude à 10, 50 et 90%.

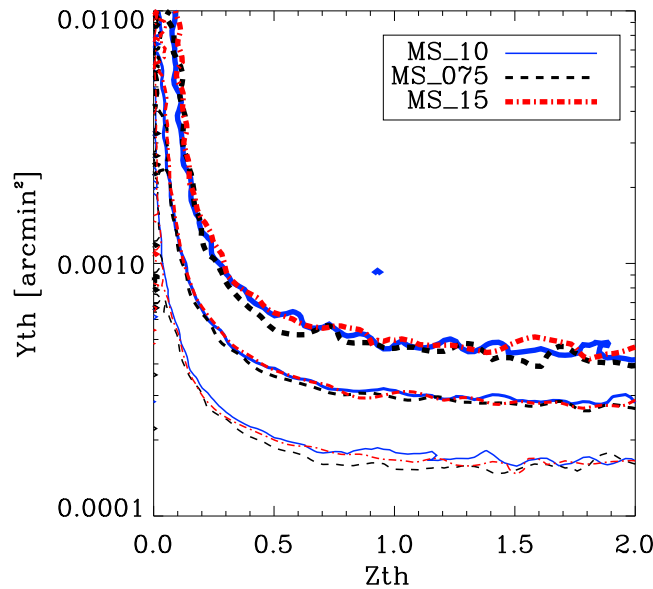


FIG. 6.3.3 – Complétude à 10, 50 et 90% (trait fin, moyen et large) pour les trois modèles statistiques *MS_10*, *MS_075* et *MS_15*.

Erreurs photométriques. La figure 6.3.4 illustre les erreurs photométriques attendues pour ces trois générations. On constate que l'erreur systématique sur l'estimation du flux des amas est indépendante de la confusion pour des sources dont le flux est supérieur à $3 \times 10^{-4} \text{ arcmin}^2$. L'erreur statistique est plus importante lorsque la confusion augmente et reste constante quand la confusion diminue.

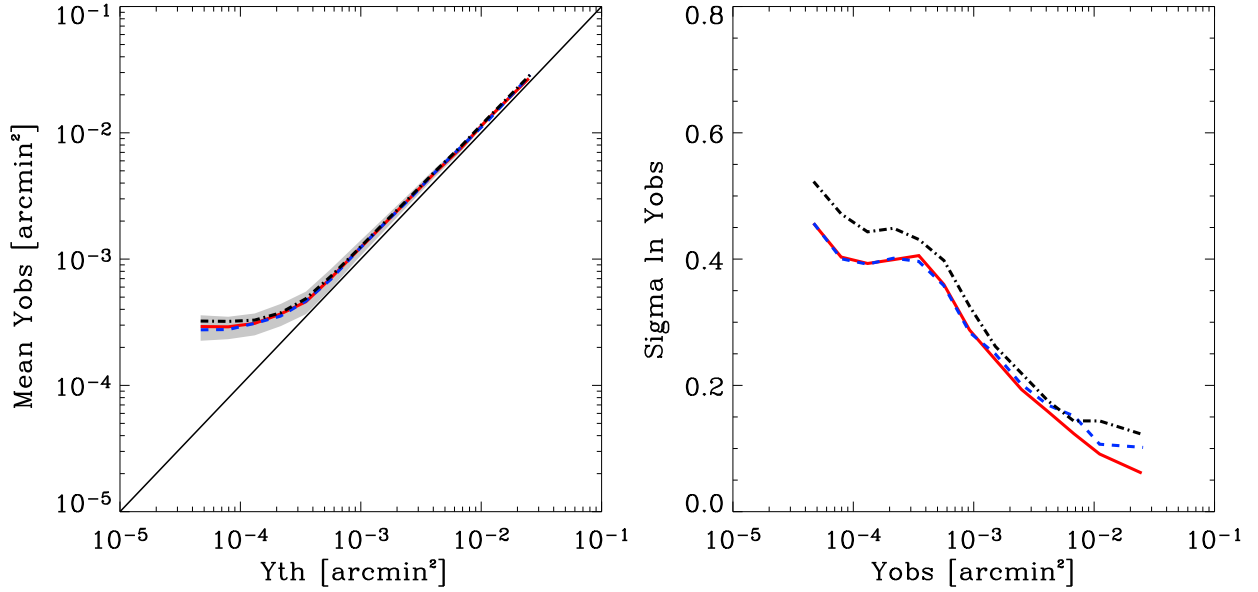


FIG. 6.3.4 – (gauche) Biais sur la reconstruction des flux pour les trois niveaux de confusion considérés. (droite) Largeur à un sigma de la distribution du logarithme du flux reconstruit. Le trait continu rouge correspond au modèle *MS_10* le trait en tirets bleu correspond au modèle *MS_075* et le trait point-tirêt noir au modèle *MS_15*.

Distribution des contaminations. La figure 6.3.5 illustre les différentes densités de probabilité des contaminations pour ces trois générations. On donne dans le paragraphe suivant la distribution du nombre total des détections pour les différents niveaux de confusion. On constate que les contaminations suivent, à un sigma, la même distribution. La différence provient essentiellement du nombre total de contamination, supposé constant et qui ne l'est pas en réalité, comme c'est montré dans le paragraphe suivant.

Nombre effectif de détections. On cherche à voir, ici, si l'écart-type des distributions : 1) du nombre d'amas, 2) du nombre de contaminations, suit bien la loi :

$$\sigma = A \sqrt{N} \quad (6.21)$$

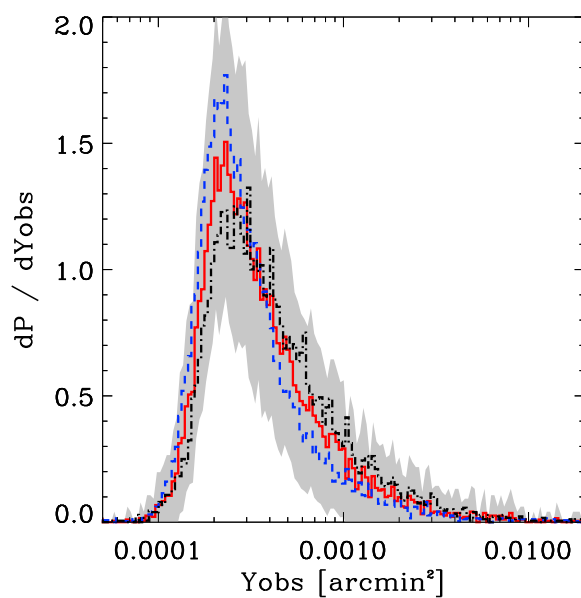


FIG. 6.3.5 – Densité de probabilité des contaminations. Le trait continu rouge correspond au modèle *MS_10* le trait en tirets bleu correspond au modèle *MS_075* et le trait point-tiret noir au modèle *MS_15*. La zone grisée donne l'erreur à un sigma pour une seule génération du Monte-Carlo.

Les valeurs moyennes étant données, pour les amas, par la complétude du sondage et pour les contaminations, supposées constantes (pour l'ensemble des détections, c'est simplement la somme des deux).

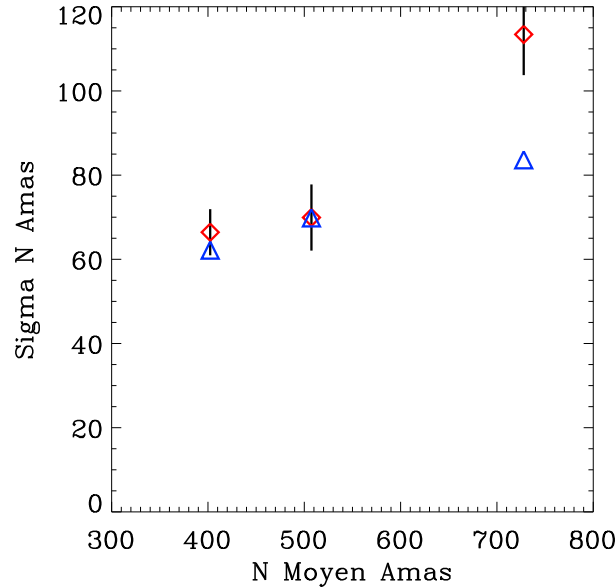


FIG. 6.3.6 – Écart-type de la distribution du nombre effectif d'amas en fonction du nombre moyen d'amas détecté. Les losanges rouges donnent les points calculés par le Monte-Carlo, les triangles bleus les points du modèle statistique. Les barres d'erreur sont données à deux sigmas.

La figure 6.3.6 donne, pour les amas, la dépendance de σ avec \bar{N} . On constate que la loi : $\sigma = A\sqrt{\bar{N}}$ avec A constant n'est pas correctement satisfaite, notamment à confusion élevée. La figure 6.3.7 donne le nombre moyen de contaminations en fonction du nombre d'amas simulés. On constate que ce nombre est quasiment constant (à deux sigmas pour les modèles *MS_075* et *MS_10*). On constate par ailleurs que la loi $\sigma = A\sqrt{\bar{N}}$ est dans les barres d'erreur uniquement à 3 sigmas.

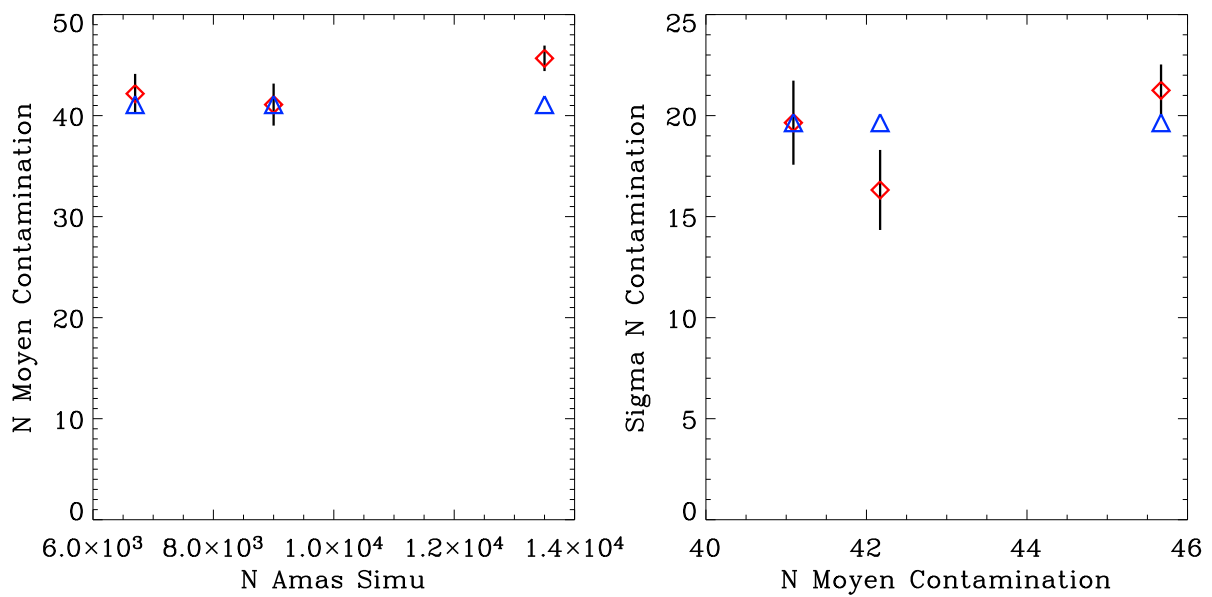


FIG. 6.3.7 – (gauche) Nombre moyen de contaminations en fonction du nombre d’amas simulés. (droite) Écart-type de la distribution du nombre effectif de contaminations en fonction du nombre moyen de contaminations. Les losanges rouges donnent les points calculés par le Monte-Carlo, les triangles bleus les points du modèle statistique. Les barres d’erreur sont données à deux sigmas.

6.4 Conclusion

Les résultats du Monte-Carlo. Le calcul par simulation Monte-Carlo de la fonction de sélection effective des sondages SZ grand-champ est un résultat important de mon travail. Les figures 6.2.2 et 6.2.3 montrent la complétude de tels sondages en fonction du redshift, du flux SZ des amas et de leur masse. L'approximation habituelle de sélection en masse n'est pas satisfaisante en pratique. Une limite en flux se rapproche des résultats du Monte-Carlo pour les amas lointains ($z > 0.8$). Cependant elle diffère, d'une part par la transition "douce" entre une complétude nulle (non-détection) et une complétude parfaite (100% de détection), d'autre part, par la détectabilité des amas proches ($z < 0.8$). La compréhension de la fonction de sélection dans cette gamme de redshift est importante car c'est là que les modèles cosmologiques standards (de type Λ CDM) prédisent un maximum d'amas massifs. On remarque par ailleurs que la détectabilité pour les redshifts inférieurs à 0.3 est plus sélective : seuls les amas les plus brillants sont détectés. Ceci est dû au fait que la brillance de surface, pour des amas de même flux total, à un redshift de 1 est beaucoup plus grande que pour un amas à un redshift de 0.3. C'est un simple effet de dilution dû à la taille apparente de l'amas qui augmente rapidement à bas redshift. Les amas proches, trop "dilués", passent sous le niveau de bruit et ne sont plus détectés.

On remarque sur la figure 6.2.5 le biais de Malmquist-Eddington dans la reconstruction du flux des amas de bas flux. Ce biais est inévitable car intrinsèque à la photométrie de sources de faible flux dans un champ bruité. En terme de distribution des sources ce biais se traduit par une surestimation des sources de faible flux. Les erreurs statistiques importantes présentes à bas flux sont probablement à l'origine de la largeur excessive de la variance des comptages. La figure 6.2.9 illustre ce point. En effet, d'une génération à l'autre, les erreurs statistiques sur les amas de faible flux conduisent à une grande variation dans le nombre total de détections, ce qui se traduit par la largeur excessive de la distribution du nombre de détections. On peut imaginer qu'en sélectionnant les amas de haut flux dans les catalogues reconstruits on limitera cet effet.

La figure 6.2.8 montre la distribution des contaminations dans les catalogues reconstruits. On constate que les contaminations ne sont pas prédominantes dans le catalogue reconstruit. Ce point est essentiel et confirme notre choix algorithmique de séparation en composante pour isoler l'effet SZ des autres émissions astrophysiques. Bien que je n'ai pas cherché à connaître l'origine des contaminants (sources infrarouges ou bruit résiduel) de manière explicite (par association au catalogue des sources infrarouges simulé) on constate que cette distribution en flux des contaminants présente une non gaussianité laissant supposer que certaines contaminations sont probablement des sources infrarouges ponctuelles. En conclusion du manuscrit je présente un travail préliminaire concernant l'expérience APEX. Il apparaît alors que les sources infrarouges sont une source importante de contamination. De plus, les sources radios, non simulées ici, seront aussi une source de contamination importante pour les cartes à basse fréquence (143 GHz). La compréhension et le calcul systématique de cette distribution est essentiel pour exploiter les catalogues d'amas. On peut imaginer que des algorithmes plus spécifiques permettant de distinguer les sources ponctuelles (infrarouges ou radios) des sources étendues que sont les amas

seront essentiels pour analyser les données d'Olimpo, APEX ou Planck (par exemple des critères de forme).

Le modèle statistique. Nous avons décrit dans ce chapitre le modèle statistique de l'expérience Olimpo (chapitres 3 et 4) et de la chaîne d'analyse et d'extraction d'amas (chapitre 5).

Le premier objectif dans la construction d'un tel modèle est d'identifier et de comprendre les effets systématiques relatifs à la détection des amas dans les observations SZ grand-champ, et à la cosmologie. Le second objectif est d'éviter l'utilisation intensive de cycle Monte-Carlo qui, en pratique, sont extrêmement gênante pour l'analyse des données et limitent les itérations possibles. Ce modèle repose sur des hypothèses fortes. J'ai montré (et je confirme au chapitre 7) qu'elles sont justifiées à la précision des prochains sondages.

Ce modèle statistique permet, à partir de la distribution théorique des amas de galaxies, de calculer la distribution de ceux qui sont effectivement observés. Ce modèle statistique de chaîne de détection SZ (instrument et analyse des données) prend en compte :

1. la complétude du sondage,
2. les erreurs photométriques d'estimation du paramètre de Compton intégré et du redshift des amas détectés,
3. les contaminations présentes parmi les détections,
4. les distributions du nombre effectif des contaminations et amas pour une observation.

Nous avons mis en évidence ses qualités et ses défauts dans la section 6.3 et esquissé les voies possibles pour l'améliorer. Afin de valider le modèle statistique comme étant un outil pour contraindre les valeurs des paramètres cosmologiques nous présentons, à la section 7.4.2, les biais et erreurs sur l'estimation des contraintes en supposant que le modèle statistique est indépendant de la confusion. Seule l'estimation des erreurs sur les paramètres cosmologiques permet de confirmer, ou non, la validité de cette hypothèse. En effet, les différences entre modèles statistiques calculés pour des niveaux de confusion différents correspondent à des variations importantes des paramètres cosmologiques, par exemple, σ_8 variant de 0.75 à 0.95. Nous répondons dans le chapitre 7 à la question de savoir si le modèle statistique est valide dans le voisinage (lequel?) du modèle cosmologique ayant servi à le générer? Dans le chapitre 7 nous avons fait l'hypothèse que le modèle statistique d'observation est indépendant des modèles (cosmologiques, de formation de structure et de physique interne des amas) ayant servi à le construire. À défaut il sera toujours possible de calculer, par Monte-Carlo, le modèle statistique de l'instrument pour un ensemble représentatif de modèles cosmologiques et, 1) soit d'interpoler les valeurs du modèle statistique pour les modèles cosmologiques manquant, 2) soit de construire des ajustements en fonction, par exemple, du nombre d'amas observés dans le champ du sondage, qui est une observable du sondage.

Contraintes sur les paramètres cosmologiques

Sommaire

7.1	Observations et cosmologie	166
7.2	Outils statistiques d'analyse des données	168
7.2.1	Compatibilité modèle-observation : les tests d'hypothèses	169
7.2.1.1	Interêt	169
7.2.1.2	L'estimateur de vraisemblance	169
7.2.1.3	Test d'hypothèse en pratique	170
7.2.1.4	Rejeter Press & Schechter	171
7.2.2	Estimation des paramètres	172
7.2.2.1	Cartes de vraisemblance	173
7.2.2.2	Calcul des niveaux de confiance par Monte-Carlo	173
7.2.2.3	L'estimateur de vraisemblance	175
7.3	Perspectives pour Olimpo	178
7.3.1	SZ, CMB et <i>weak-shear</i> : contraintes sur T_\star	179
7.3.2	Ω_M et σ_8	181
7.3.3	Variance des comptages et contraintes	181
7.3.4	Suivi optique partiel	184
7.3.5	Contraintes "robustes"	184
7.4	Systématiques liées au modèle statistique	187
7.4.1	Estimateur des paramètres cosmologiques	187
7.4.2	Modèle statistique et confusion	188
7.4.3	Complétude	191
7.4.4	Biais photométrique	191
7.4.5	Contamination	192
7.5	Conclusion	194

7.1 Observations et cosmologie

DANS le chapitre précédent j'ai présenté notre modèle statistique de chaîne de détection SZ. Ce modèle permet de calculer, à partir de la distribution théorique en flux et redshift des amas, la distribution des détections effectivement observées :

$$\frac{dN_{th}}{dz_{th}dY_{th}}(z_{th}, Y_{th}) \longrightarrow \text{Modèle statistique} \longrightarrow \frac{dN_{obs}}{dz_{obs}dY_{obs}}(z_{obs}, Y_{obs})$$

avec z_{obs} et Y_{obs} les observables mesurées dans l'expérience et z_{th} et Y_{th} les flux et redshift "vrais" de ces détections. La distribution des détections calculée inclut : la complétude du sondage, les erreurs photométriques en flux et redshift, les contaminations et les distributions du nombre effectif de sources détectées (amas et contamination).

La figure 7.1.1 illustre le travail de l'analyse statistique des données observées. On y voit que la construction du modèle statistique de la chaîne de détection est es-

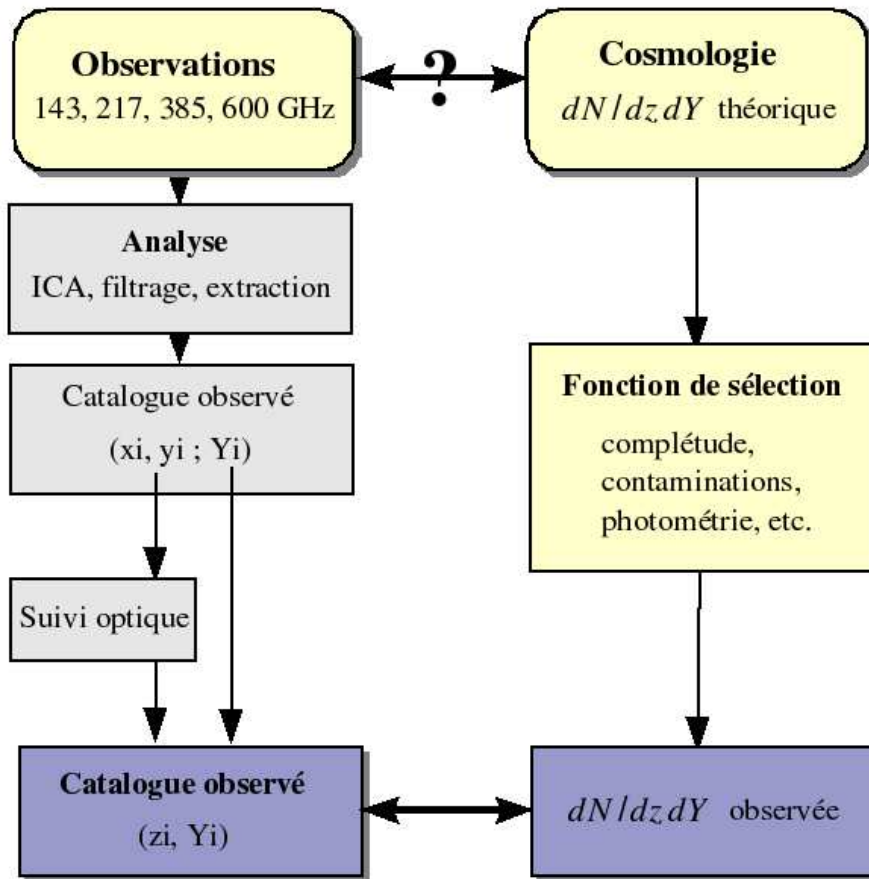


FIG. 7.1.1 – Schéma illustrant le travail d'analyse statistique d'une observation.

sentielle pour estimer correctement les contraintes sur les paramètres cosmologiques.

Le travail caché derrière la double flèche au bas de cette figure est exposé à la section 7.2 de ce chapitre. La partie en bleu, à gauche, illustre l'analyse des données, et la partie droite illustre l'utilisation du modèle statistique de cette même chaîne.

Dans ce chapitre je montre quel sera l'apport des sondages SZ d'Olimpo en termes de contraintes sur les paramètres cosmologiques Ω_M et σ_8 . J'illustrerai ce résultat par des calculs de contraintes pour différents types de suivi optique. Pour cela je présente, dans un premier temps, les outils statistiques que j'ai mis en place pour réaliser ces études : tests d'hypothèses et calcul de contraintes. Ensuite j'étudie certains effets systématiques, en concluant sur la validité du modèle statistique de chaîne de détection SZ présenté au chapitre 6.

7.2 Outils statistiques d'analyse des données

L'analyse de données d'expériences de type Olimpo en vue de contraindre les paramètres des modèles cosmologiques consiste à comparer la distribution attendue des amas observés, en redshift et flux, pour un modèle cosmologique donné avec la distribution effectivement observée (i. e., le catalogue (z_i, Y_i) , voir le schéma de la figure 7.1.1). Le modèle statistique de la chaîne de détection prend ici toute son importance : il permet, en effet, de calculer rapidement (à la différence d'une génération complète Monte-Carlo) la distribution théorique observée des amas pour n'importe quel modèle cosmologique. La figure 7.2.1 illustre les deux “objets” à comparer : le catalogue observé et la distribution observée calculée théoriquement.

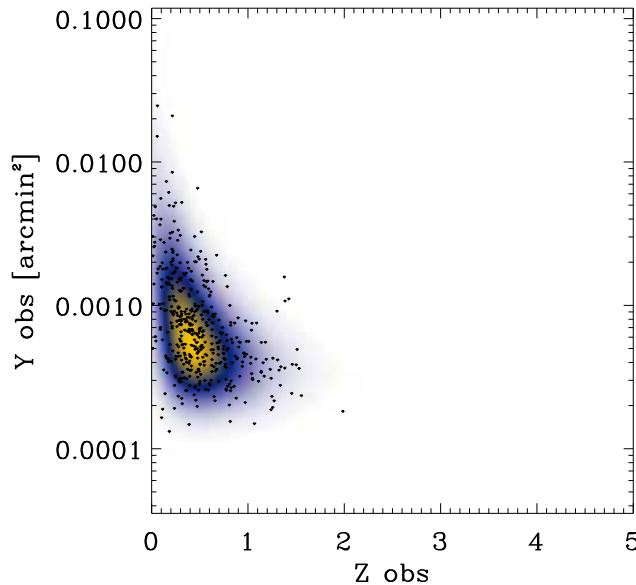


FIG. 7.2.1 – Distributions, en flux et redshift, calculée (surface de densité) et observée (points (z_{obs}, Y_{obs}) du catalogue).

Dans la suite du chapitre les contours sont donnés à 1, 2 et 3 σ . Pour les zones oranges (grises) le contour à 1 σ est le plus foncé. Pour les courbes de niveau le contour à 1 σ est donné en trait plein. Les courbes de niveau donnent, pour toutes les figures, les contours de référence en vue de comparaison avec un cas particulier donné par les zones oranges (grises). Le losange blanc indique le modèle de concordance donné dans le tableau 6.1. La croix grise (orangée) donne le meilleur modèle reconstruit pour les zones grisées (orangées) tandis que la croix noire donne le meilleur modèle reconstruit pour les contours de référence en courbes de niveaux noires.

7.2.1 Compatibilité modèle-observation : les tests d'hypothèses

Le premier outil statistique que nous avons mis en place pour l'étude d'une observation grand champ d'amas SZ est le test d'hypothèse. Nous avons développé dans un premier temps un test de Kolmogorov-Smirnov. Mais son utilisation en pratique s'est révélée trop coûteuse en temps de calcul et nous avons ensuite mis en place un test d'hypothèse basé sur l'estimateur de vraisemblance étendue L_E (nommé dans la suite simplement vraisemblance, L pour *Likelihood*). Le principe est le même pour les deux tests. La différence provient uniquement de la manière dont on mesure "l'écart" entre la distribution des observations et la distribution théorique prédite par le modèle cosmologique. Cet outil statistique est complémentaire de l'outil de calcul des paramètres cosmologiques.

7.2.1.1 Interêt

Pour chacun des modèles cosmologiques \vec{P} que l'on veut tester on fait l'hypothèse suivante : "l'observation considérée est une génération aléatoire du modèle \vec{P} ". Les tests d'hypothèses permettent de calculer la probabilité que cette hypothèse soit vraie. Plusieurs arguments ont guidé notre choix vers cet outil statistique : tout d'abord le fait que les prochaines expériences Olimpo ou APEX auront, dans un premier temps, des données de "test" (vol test pour Olimpo et premiers mois de prise de donnée pour APEX) comptant moins d'une centaine d'amas, ensuite le fait que les modèles de la physique interne des amas soient peu contraints. Les tests d'hypothèses permettent de trouver une classe de modèles compatibles avec les données. On peut imaginer placer des contraintes sur les paramètres cosmologiques sans faire ce travail préalable, et aboutir à des contraintes raisonnables alors que les modèles testés ne sont pas statistiquement compatibles avec les données. En pratique l'analyse statistique des données se compose des trois étapes suivantes : 1) on conduit un ensemble de test d'hypothèses sur une large gamme de modèles, 2) on sélectionne ceux qui sont statistiquement compatibles avec nos données, 3) on cherche le modèle le plus en accord avec les données et on calcule les contraintes sur les paramètres testés (i.e. les erreurs sur l'estimation du meilleur modèle).

Par exemple, un test d'hypothèse permet de décider si un modèle cosmologique \vec{P} qui serait en accord avec d'autres observations indépendantes (CMB, supernovae, amas en X, etc.) est incompatible avec nos données. On est en mesure de dire si l'hypothèse "mon observation est une génération aléatoire de ce modèle \vec{P} " est vraie ou fausse, et avec quelle probabilité.

7.2.1.2 L'estimateur de vraisemblance

L'estimateur de vraisemblance est défini d'une manière générale par :

$$L = \prod_{i=1}^N f(x_i; \theta) \quad (7.1)$$

où les variables x_i sont les N mesures et f la p.d.f théorique du modèle testé fonction du paramètre θ .

Dans notre cas nous devons tenir compte des contaminations (les sources que le suivi optique indique comme n'étant pas des amas), des amas dont on a mesuré le redshift (on est alors sûr que ce sont des amas) et des sources dont on ignore si ce sont des amas ou des contaminations (pas de mesure en redshift). L'estimateur de vraisemblance L que nous utilisons pour comparer une observation exp et un modèle cosmologique \vec{P} s'écrit alors de la manière suivante :

$$L(\vec{P}|exp) = \frac{dP}{dN_{det}}(N_{exp}^{det}, \vec{P}) \prod_{i=1}^{N_{exp}^{amas}} \frac{dP}{dz_{obs}dY_{obs}}(z_i^{amas}, Y_i^{amas}; \vec{P}) \prod_{i=1}^{N_{exp}^{detection}} \frac{dP}{dY_{obs}}(Y_i^{detection}; \vec{P}) \quad (7.2)$$

Le premier terme du produit permet d'estimer l'information contenue dans le nombre total de détections. Le nombre de sources détectées est une information en soi, indépendante des mesures individuelles de flux ou redshift. La p.d.f est donnée par le modèle statistique et N_{exp}^{det} est le nombre total de sources brillantes détectées dans le champ. Le premier produit permet de mesurer l'information contenue dans les amas détectés (dont on est sûr que ce sont des amas grâce au suivi optique) étant donnés leurs redshift z_i^{amas} et flux Y_i^{amas} . Le dernier produit mesure l'information contenue dans le flux $Y_i^{detection}$ des détections dont on ne sait pas si elles sont des contaminations ou des amas. La p.d.f correspondante est calculée à partir de la distribution en flux des amas à laquelle on a ajouté la distribution en flux des contaminations. Les détections dont on est sûr que ce sont des contaminations n'apportent aucune information sur la cosmologie. Elles sont exclues du calcul de la vraisemblance.

En pratique on travaille avec les quantités $-\ln L$. Les méthodes de "maximum de vraisemblance" deviennent alors des méthodes de "minimum d'in-vraisemblance".

Le modèle statistique joue ici un rôle primordial dans le calcul des paramètres cosmologiques à travers la connaissance des différentes p.d.f observées-simulées. Dans le paragraphe 7.4 ses limites en tant qu'outil "d'estimation" des paramètres cosmologiques ont été démontrées.

7.2.1.3 Test d'hypothèse en pratique

Les étapes pour effectuer un test d'hypothèse sur un modèle cosmologique vis-à-vis d'une observation sont en pratique les suivantes :

1. on dispose d'une observation,
2. on se place dans le cadre d'un modèle cosmologique \vec{P} que l'on désire tester,
3. on calcule la distribution théorique des amas dans le cadre de ce modèle,
4. le modèle statistique, construit au préalable, nous donne les distributions effectivement observées des amas et des contaminations (typiquement : leur nombre, la dispersion sur ce nombre, les distributions en fonction des observables),
5. on génère par simulation de Monte-Carlo un ensemble d'observations simulées compatibles avec ce modèle \vec{P} ,
6. chacune à une vraisemblance vis-à-vis du modèle testé $L_i = L(\vec{P}|exp_i)$. On construit la distribution de ces vraisemblances,

7. à partir de cette p.d.f de vraisemblance (7.2.2 gauche) on peut construire la courbe de probabilité de compatibilité (7.2.2 droite),
8. la vraisemblance de l'observation que l'on analyse vis-à-vis du modèle testé correspond alors à une probabilité dans cette courbe : c'est la probabilité de compatibilité recherchée.

La construction de la courbe de probabilité se fait en calculant, dans un premier temps la valeur de la vraisemblance de référence L_0 : le pic de la distribution. Celle-ci étant bien ajustée par une gaussienne, on considère comme référence la moyenne de la distribution. La probabilité de compatibilité de l'observation est donnée par le rapport du nombre de générations ayant un écart de vraisemblance supérieur à l'observation exp sur le nombre total de génération N :

$$Prob_{exp} = \frac{N(\Delta L_i > \Delta L_{exp})}{N} \quad (7.3)$$

La figure 7.2.2 illustre les trois dernières étapes.

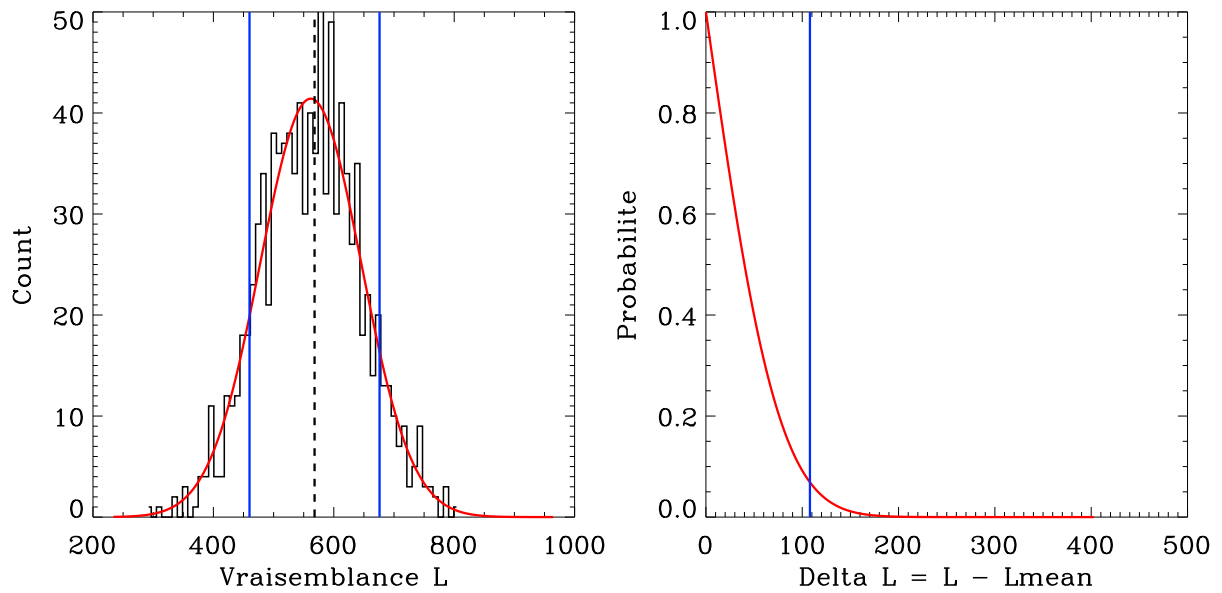


FIG. 7.2.2 – (gauche) Distribution des vraisemblances pour N générations aléatoires d'observations à partir du modèle cosmologique à tester. (droite) Courbe de probabilité déduite de la p.d.f. Les traits verticaux bleus donnent la valeur de la vraisemblance de l'observation. La courbe rouge indique l'ajustement gaussien effectué sur les données du Monte-Carlo représentées par l'histogramme noir. Le trait vertical discontinu indique la valeur de L_0 .

7.2.1.4 Rejeter Press & Schechter

Les tests d'hypothèses peuvent permettre, par exemple, de calculer la probabilité de compatibilité entre une observation, simulée dans notre cas avec la fonction de

masse de Sheth et Tormen (et pour le modèle cosmologique de concordance), et un modèle cosmologique dans lequel la distribution des amas est calculée grâce à la fonction de masse de Press et Schechter.

On fait l'hypothèse suivante : "l'observation est une génération aléatoire d'une distribution d'amas calculée grâce à la fonction de masse de Press & Schechter". On calcul, en utilisant la méthode précédente, la courbe de probabilité en fonction des différences de vraisemblance, pour ce modèle physique de formation d'amas. Elle est donnée dans la figure 7.2.3. La probabilité de compatibilité pour l'observation considérée est inférieure à 1×10^{-5} . Avec les données du vol scientifique d'Olimpo on devrait être en mesure de tester les différentes fonction de masse (Sheth & Tormen, Press & Schechter, Jenkins & al.) en utilisant les paramètres cosmologiques estimés par d'autres expériences.

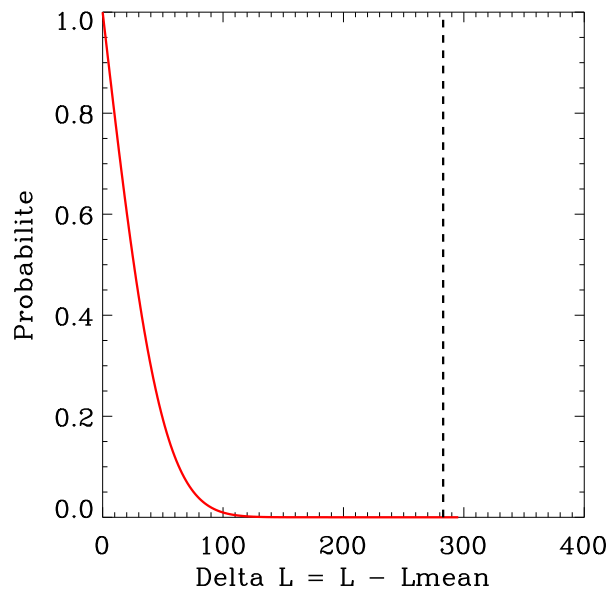


FIG. 7.2.3 – En trait continu la courbe de probabilité pour le modèle de concordance avec utilisation de la fonction de masse de Press & Schechter. En trait discontinu la valeur de la vraisemblance de l'observation considérée (calculée avec la fonction de masse de Sheth & Tormen). L'observation a une probabilité d'être une génération aléatoire du modèle cosmologique considérée inférieure à 1×10^{-5} . Elle est donc incompatible à 5σ avec le modèle théorique.

7.2.2 Estimation des paramètres

Grâce au test d'hypothèse présenté dans la section précédente on est en mesure de préciser quelle classe de modèles est compatible avec une observation. La seconde étape de l'analyse statistique des données est l'estimation, parmi les modèles compatibles, du modèle le mieux "en accord" avec les données, de donner les paramètres

de ce modèle et d'établir les erreurs statistiques sur ces paramètres. Le test d'hypothèse donne la probabilité qu'une observation soit une génération aléatoire d'un modèle; l'estimation des contraintes donne, pour une observation, les erreurs faites si on considère un modèle voisin du modèle le plus en accord avec celle-ci.

La démarche complète de calcul des contraintes est :

1. on dispose d'une observation,
2. on trouve la classe des modèles compatibles avec cette observation par des tests d'hypothèses,
3. parmi ceux-ci on cherche "le mieux en accord" (en termes de vraisemblance),
4. on calcule les niveaux de confiance (i.e., les erreurs statistiques) autour de ce meilleur modèle.

7.2.2.1 Cartes de vraisemblance

En pratique pour calculer les contraintes on construit une "carte" de vraisemblance pour un ensemble de modèles \vec{P} et l'observation considérée (dénotée exp dans la suite). La maximisation de la vraisemblance désigne le modèle \vec{P}^{min} le plus en accord avec les données (qui a, par rapport aux données d'observation exp , une vraisemblance $L^{min}(\vec{P}^{min}|exp)$ maximum). En pratique on parcourt les paramètres des modèles \vec{P} et on cherche le minimum de la quantité :

$$-\ln L(\vec{P}|exp) \quad (7.4)$$

Les niveaux de confiance sont calculés par rapport au minimum de la vraisemblance L^{min} . On construit la carte de la différence de vraisemblance :

$$\Delta L(\vec{P}|exp) = -2 \left[\ln L(\vec{P}|exp) - \ln L^{min}(\vec{P}^{min}|exp) \right] \quad (7.5)$$

Cette quantité, dans le cas où les vraisemblances sont calculées à partir de p.d.f gaussiennes, suit une loi de χ^2 . En pratique on calcule, par Monte-Carlo, les distributions de "différence de vraisemblance" (appelée dans la suite : log-vraisemblance) afin de vérifier cette hypothèse (voir le paragraphe suivant). Si, dans un premier temps, on suppose que les log-vraisemblances entre un modèle \vec{P} et le meilleur modèle \vec{P}^{min} suivent une loi de χ^2 alors, à partir de la carte des log-vraisemblances, on peut calculer directement les niveaux de confiance autour du meilleur modèle. La figure 7.2.4 illustre les différentes étapes de cette démarche.

7.2.2.2 Calcul des niveaux de confiance par Monte-Carlo

La démarche pour déterminer, par Monte-Carlo, les niveaux de confiance est la suivante :

1. on génère, en accord avec le modèle \vec{P}^{min} , N observations simulées,
2. pour chacune d'entre elles on cherche, par maximisation de la vraisemblance, le modèle de meilleur accord \vec{P}^i ,

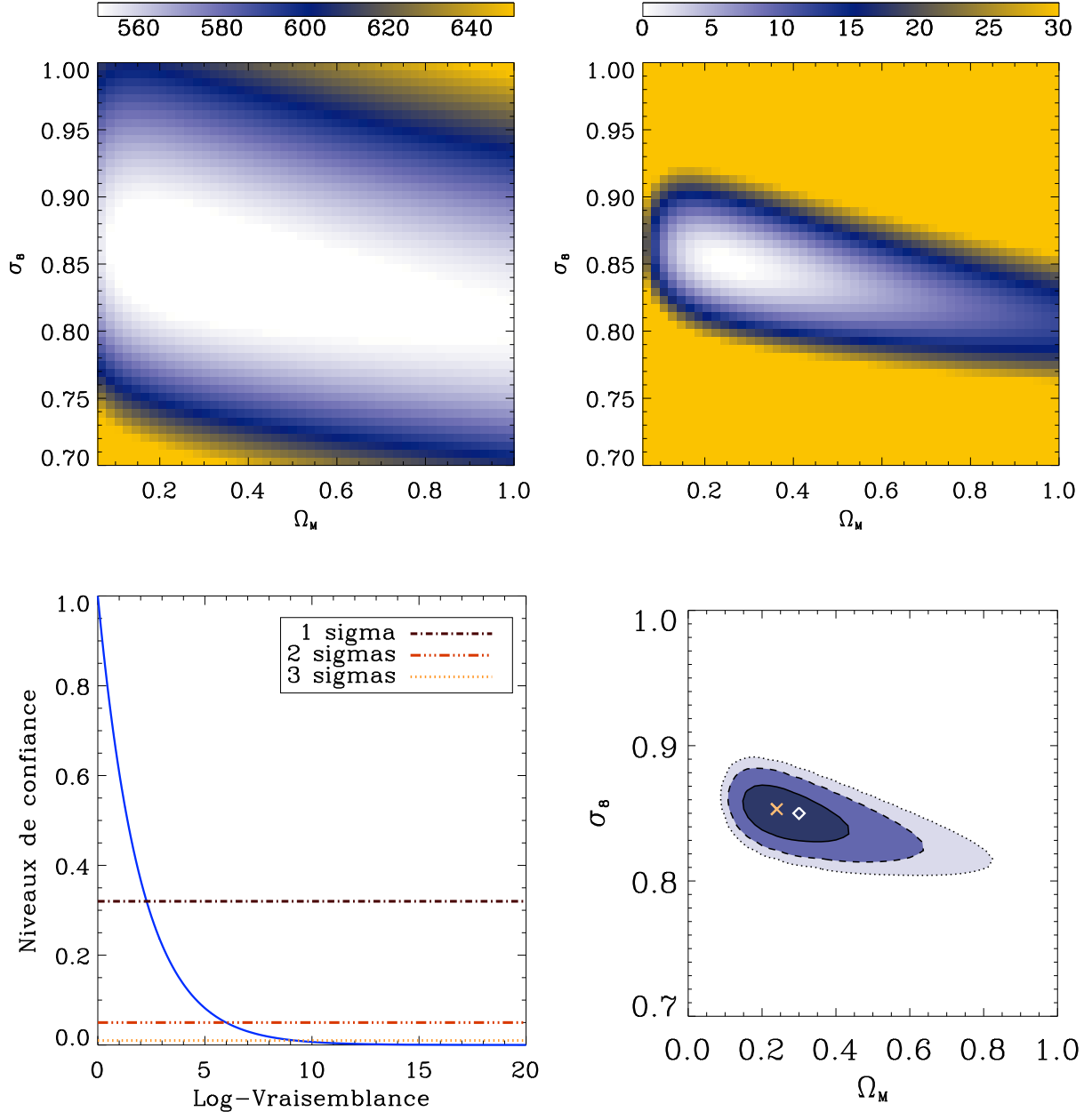


FIG. 7.2.4 – De gauche à droite et de haut en bas : 1) la carte de vraisemblance des modèles $\vec{P} = (\Omega_M, \sigma_8)$, 2) la carte de log-vraisemblance, 3) la distribution en χ^2 des log-vraisemblances (les traits horizontaux donnent les niveaux de confiance à 1, 2 et 3 sigmas), 4) les niveaux de confiance calculés grâce à cette distribution et à la carte de log-vraisemblance au dessus (donnés à 1, 2 et 3 sigmas).

3. pour chacun des N modèle \vec{P}^i on calcule la vraisemblance L^i vis-à-vis de l'observation initiale *exp*,
4. toutes les vraisemblances calculées sont, par construction de \vec{P}^{min} , inférieures (moins “en accord”) à la vraisemblance entre l'observation *exp* et le modèle \vec{P}^{min} ,
5. la distribution des différences $\Delta L^i = -2[\ln L^i(\vec{P}) - \ln L^{min}(\vec{P}^{min})]$ donne les niveaux de confiance que l'on peut appliquer, ensuite, à la carte de ΔL précédente.

La figure 7.2.5 illustre ces étapes.

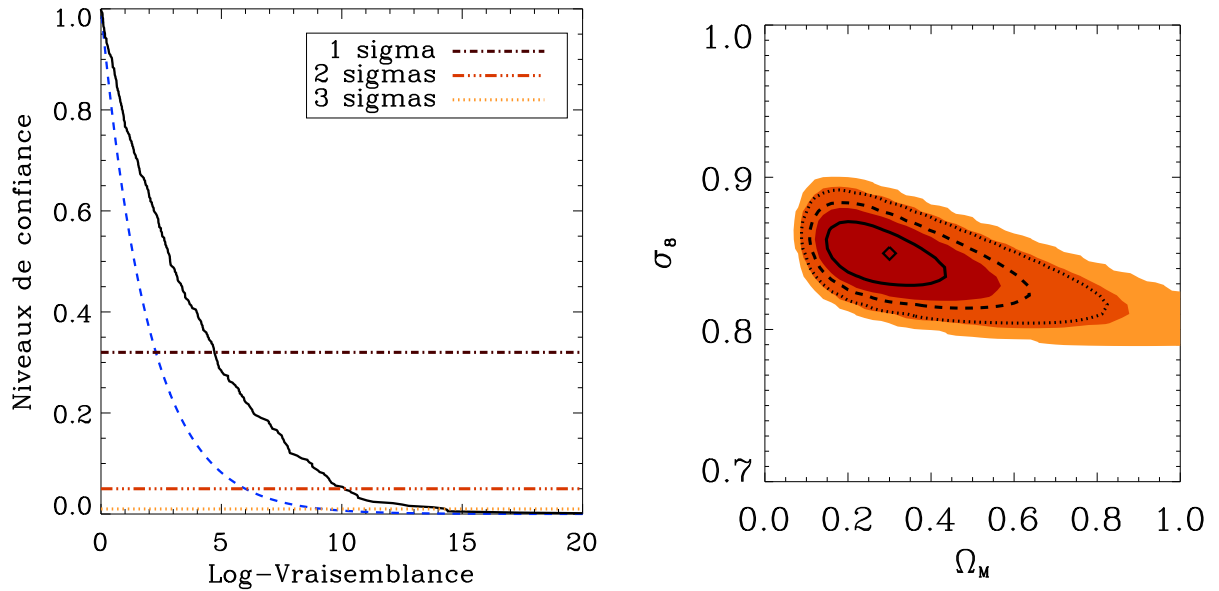


FIG. 7.2.5 – (gauche) Distribution des différences de vraisemblance (log-vraisemblance) calculée par Monte-Carlo (courbe noire), superposée à la distribution théorique de χ^2 (courbe bleue). (droite) Impact sur les contraintes de la distribution calculée par Monte-Carlo. Les régions oranges correspondent aux contraintes à 1, 2 et 3 sigmas pour la distribution calculée par Monte-Carlo tandis que les contours superposés donnent les résultats si on suppose une distribution de χ^2 (trait continu : 1 sigma, trait discontinu : 2 sigmas, pointillés : 3 sigmas).

7.2.2.3 L'estimateur de vraisemblance

La figure 7.2.6 illustre les influences de la mesure du nombre de détections d'une part et de la “forme” de la distribution en redshift et flux d'autre part. Pour 500 détections, dont on mesure le redshift, l'information sur la forme domine l'information sur le nombre. Pour moins de 80 détections (vol test d'Olimpo) l'information

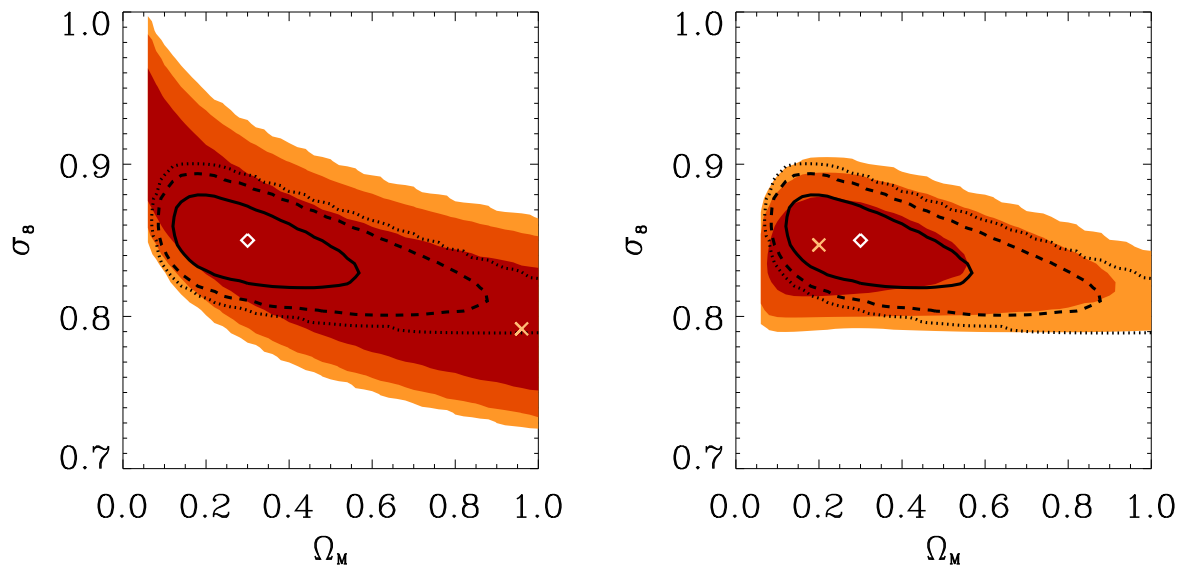


FIG. 7.2.6 – (gauche) Paramètres cosmologiques calculés grâce à la mesure du seul nombre de détections. (droite) Paramètres cosmologiques calculés grâce à la seule forme de la distribution en flux et redshift. En superposition, les contours obtenus en cumulant les deux informations. Le code de couleur des contours est le même pour tout le chapitre (voir la légende 7.2.5).

sur le nombre domine et les paramètres Ω_M et σ_8 deviennent rapidement dégénérés.

Il est intéressant de remarquer que les contours donnés par l'information sur le total de détections sont relativement lâches. Ceci est dû au fait que les distributions effectives du nombre de détections ne suivent pas loi de poisson mais une loi gaussienne dont l'écart-type est de l'ordre de : $3 \times \sqrt{N}$. En première approximation on considère souvent une distribution de poisson, ce qui conduit à surestimer la qualité des contraintes données par le nombre de détections.

7.3 Perspectives pour Olimpo

Les buts des sondages d'amas SZ sont de comprendre la physique du gaz des amas, la formation des grandes structures et enfin les modèles cosmologiques. Ces différents modèles possèdent des paramètres et impliquent certaines hypothèses qui peuvent être testées par ces prochains sondages SZ. Dans la suite je montre comment les observations SZ permettront de comprendre la physique du gaz des amas en contraignant le paramètre T_* . Ensuite, en supposant ce paramètre bien établi, j'explore le potentiel de ces observations pour contraindre les paramètres des modèles cosmologiques Ω_M et σ_8 . Je termine en présentant quelques études d'effets systématiques.

Dans cette section je présente différents résultats illustrant les possibilités, en termes de contraintes, d'un sondage grand champ d'Olimpo. On considère un sondage de 400 deg^2 avec un vol de dix jours. Les caractéristiques instrumentales utilisées sont données dans le chapitre 3 (voir le tableau 3.1). On considère le catalogue observé présenté dans la figure 7.3.1. Ce catalogue est constitué de 519 sources brillantes

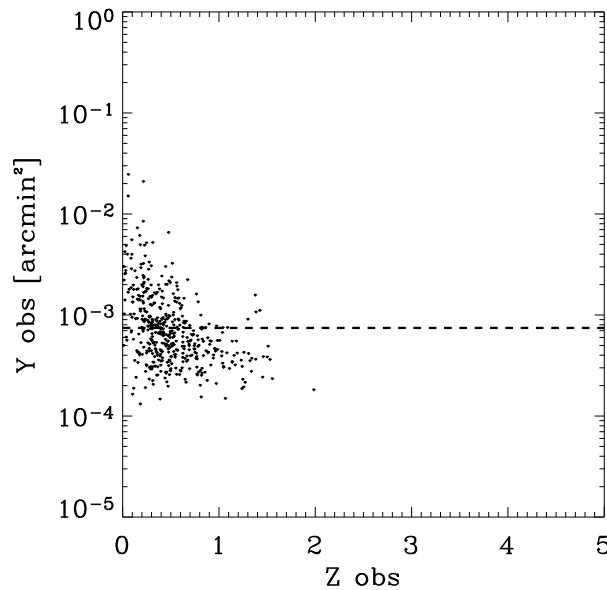


FIG. 7.3.1 – Catalogue des amas détectés d'une observation typique générée en accord avec le modèle de concordance pour un vol scientifique d'Olimpo. La ligne en tirets illustre la limite en flux considérée dans la section 7.3.5.

détectées. Parmi celle-ci, et si le suivi pointe l'ensemble des sources, on sera en mesure d'identifier 469 amas et 50 contaminations. Cette observation est une génération aléatoire du modèle de concordance dont les paramètres sont donnés dans le tableau 4.1.

7.3.1 SZ, CMB et *weak-shear* : contraintes sur T_\star

Les phénomènes galactiques chauffent le gaz intra-amas et contribuent à l'effet SZ de l'amas en plus du chauffage gravitationnel. Ces phénomènes de physique interne sont encore mal compris et leur contribution au chauffage du gaz est paramétrisée par la quantité effective T_\star de normalisation de la loi d'échelle M-T.

La forte dégénérescence entre les paramètres : de formation de structure σ_8 , et de la physique de chauffage du gaz dans les amas en équilibre : T_\star , fait qu'il est nécessaire de contraindre T_\star par ailleurs pour pouvoir exploiter les données des sondages SZ (et X, de la même manière) en termes de contraintes sur les paramètres *cosmologiques* : σ_8 et Ω_M principalement. Dans cette optique nous avons, dans un premier temps, estimé les contraintes sur T_\star que l'on peut attendre avec les données d'Olimpo combinées aux données de WMAP de mesure du spectre de puissance du CMB et aux données prochainement attendues pour les observations *weak-shear* du sondage CFHTLS-wide (170 deg² avec 2 *bins* de redshift). Le weak-shear permet une bonne mesure de σ_8 indépendante de T_\star . Par ailleurs, les données de WMAP donnent une information précise sur Ω_M .

En pratique on calcule les contraintes sur σ_8 et T_\star en marginalisant les données SZ sur le paramètre Ω_M . Cette approche constitue une analyse, au premier ordre, des contraintes que l'on peut attendre sur T_\star . La vraisemblance combinée de ces trois jeux de données est calculée en utilisant :

- la vraisemblance complète présentée dans le début du chapitre pour les données SZ,
- pour les données de WMAP et du CFHTLS : la vraisemblance approchée donnée par la matrice de Fisher sur les paramètres Ω_M et σ_8 marginalisée sur tous les autres paramètres (h , n_s , Ω_b , τ dans un cadre cosmologique de type Λ CDM).

La matrice de Fisher a été calculée par C. Yeche (CEA-Dapnia/SPP) dans le cadre d'un article soumis à A&A : Yeche et al. (2005). Le tableau 7.3.1 donne les erreurs statistiques et la corrélation sur Ω_M et σ_8 calculées pour la combinaison des données de WMAP et du CFHTLS-wide.

Ω_M	σ_8	corrélation
0.0332	0.0096	0.18

TAB. 7.1 – Erreurs à un σ sur l'estimation des paramètres Ω_M et σ_8 et corrélation entre les paramètres. Ces contraintes sont celles données par les observations combinées d'un vol scientifique d'Olimpo nominal avec un suivi optique complet, du CMB par WMAP et du *weak-shear* attendues pour le sondage CFHTLS-wide.

La figure 7.3.2 illustre ces résultats. La figure de gauche illustre la forte dégénérescence entre les deux paramètres T_\star et σ_8 si on ne considère que les données SZ. Les données du CMB et du *weak-shear* apportent des contraintes fortes sur Ω_M et σ_8 permettant de lever la dégénérescence (figure de droite). Les contraintes obtenues sur T_\star restent, cependant, assez lâches et il apparaît clairement ici, que des observations en X complémentaires permettraient d'améliorer encore les contraintes.

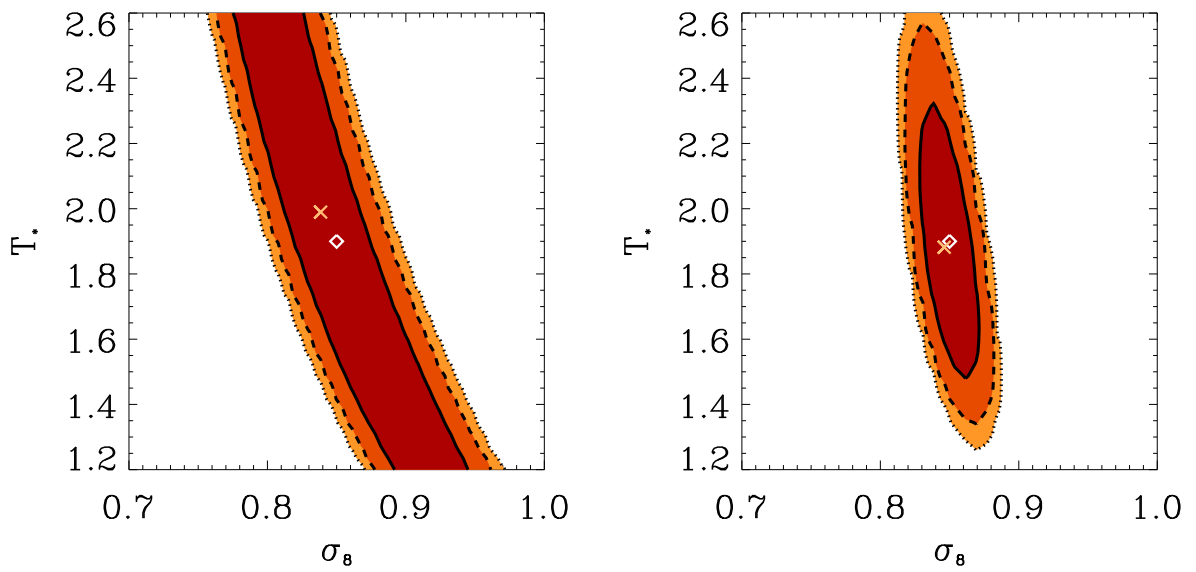


FIG. 7.3.2 – Contraintes sur les paramètres σ_8 et T_* attendues pour le vol scientifique d'Olimpo. Les contours sont calculés en marginalisant sur le paramètre Ω_M et en supposant un suivi complet des sources détectées. A gauche, les contours attendus avec le sondage SZ seul. A droite, les contours attendus avec la combinaison des données du sondage SZ et des données de WMAP et du sondage *weak-shear* du CFHTLS-wide. La croix donne la position du meilleur modèle reconstruit, le losange la position du modèle de concordance.

Dans la suite de ce chapitre, nous nous intéressons aux contraintes sur les paramètres Ω_M et σ_8 . Il apparaît, à la vue des contraintes sur T_* , que si ce paramètre est laissé libre les erreurs sur ces deux paramètres seront importantes du fait des dégénérescences. Notre objectif étant principalement d'illustrer les erreurs systématiques attendues pour les prochains sondages SZ grand-champ, nous considérons T_* fixé à la valeur du modèle de concordance. Cette approche, minimisant les erreurs statistiques, nous permettra de mettre en évidence les erreurs systématiques.

7.3.2 Ω_M et σ_8

On se limitera, pour le moment, et bien que les méthodes soient directement transposables à d'autres paramètres, à l'étude des paramètres Ω_M (avec Ω_B constant) et σ_8 . Ces deux paramètres sont sensibles aux comptages d'amas. Ω_M est principalement sensible à la distribution en redshift des amas et σ_8 aux comptages globaux et à la distribution en flux. Les autres paramètres des modèles cosmologiques et de formation de structures sont supposés connus sans erreurs. La suite de mon travail consiste à tenir compte d'information *a priori* sur la valeur de ces paramètres et leur erreur.

La figure 7.3.3 illustre les résultats attendus d'une observation typique d'Olimpo. Les caractéristiques instrumentales sont données dans le chapitre 3. La chaîne d'analyse des données et d'extraction des amas est présentée au chapitre 5. On a supposé, ici, pouvoir accéder à un suivi optique complet des 519 sources. L'influence sur l'estimation des paramètres cosmologiques d'un suivi qui ne serait que partiel est donnée dans la section suivante. Les erreurs à un σ valent, pour Ω_M : $+0.31/-0.15$, et pour σ_8 : $+0.028/-0.035$. La corrélation entre les deux paramètres étant donnée par la figure. L'erreur systématique est donnée dans la section 7.4.

7.3.3 Variance des comptages et contraintes

Comme je l'ai montré au chapitre 6 (6.2.4) la variance des comptages du nombre de sources détectées est anormalement plus grande que la variance d'une distribution de Poisson de même moyenne, valeur attendue pour N observations indépendantes. Le comptage des amas est riche d'information sur les modèles cosmologiques, celle-ci étant quantifiée par le premier terme de la vraisemblance définie par l'équation 7.2. La figure 7.3.4 montre l'impact du calcul de la variance des comptages par Monte-Carlo sur les contraintes attendues. L'impact de la variance des comptages apparaît dans cette figure de manière claire. Le calcul par Monte-Carlo induit une augmentation des erreurs statistiques sur le paramètre σ_8 de l'ordre de 100%. Ainsi il apparaît essentiel de comprendre les algorithmes de détection d'amas, non seulement sur leur efficacité et leur photométrie mais surtout sur la variance des comptages qu'ils induisent en sortie de chaîne de détection.

Je n'ai pas pu déterminer de manière précise d'où provient cette variance "excessive". Trois causes sont cependant privilégiées.

- Le premier suspect est l'algorithme de filtrage non-linéaire qui, par ses choix automatiques de seuil d'une génération à l'autre, peut induire un filtrage plus ou moins fort et donc faire fortement varier le nombre de sources détectées.

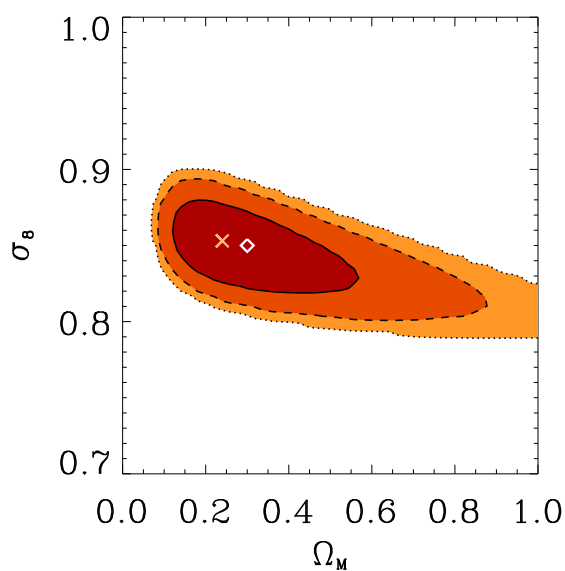


FIG. 7.3.3 – Contraintes sur les paramètres Ω_M et σ_8 attendues pour le vol scientifique d'Olimpo. Les contraintes sont calculées en supposant un suivi complet des sources détectées. Ces contraintes servent de référence dans tout ce chapitre, elles seront redonnées, en superposition, pour tous les contours.

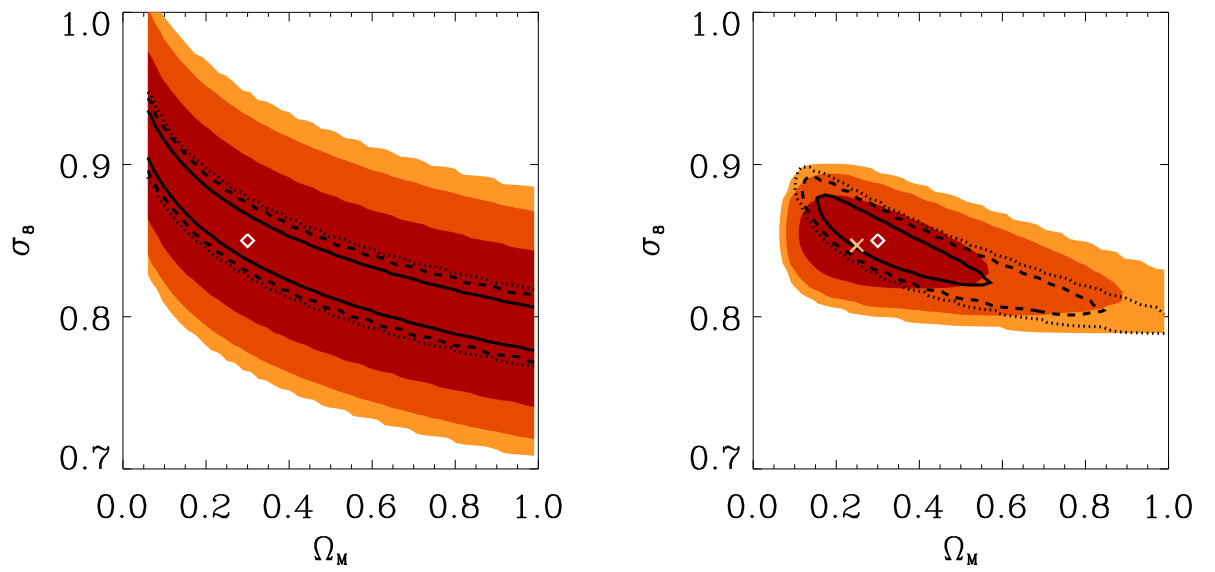


FIG. 7.3.4 – Dégradation des contraintes due à la variance des comptages. A gauche, si l'on ne prend en compte que l'information sur le comptage des amas, à droite, si toutes les informations sont prises en compte : comptage et mesures de redshift et de flux. Les niveaux colorés donnent les contraintes avec le calcul par Monte-Carlo tandis que les contours en noir donnent les contraintes obtenues en supposant une distribution de Poisson pour N comptage indépendants.

- Le second suspect est la confusion entre sources : la présence dans une génération d'un amas proche et massif a pour effet de masquer un grand nombre d'amas qui ne seront alors plus détectés.
- Le troisième suspect est l'importante erreur statistique sur la photométrie à bas flux. On s'attend à ce que le nombre d'amas de bas flux détectés varie beaucoup d'une génération à l'autre.

Ces trois pistes seront explorées dans la suite de ce travail :

- en testant un filtrage linéaire simple,
- en limitant la confusion entre amas,
- et en se limitant aux amas de haut flux pour lesquels la photométrie est meilleure.

7.3.4 Suivi optique partiel

Les sondages SZ avec caméra bolométrique vont fournir de grands catalogues de sources qui seront au moins en partie pointés par des expériences de suivi optique afin de mesurer le redshift des amas détectés.

Il est intéressant d'étudier l'influence de suivis optiques incomplet sur les paramètres (Ω_M, σ_8). On en considère trois types :

1. *COMPLETE* : 100 % des sources sont observées lors du suivi.
2. *RAN50* : 50% des sources sont observées lors du suivi. Ces sources étant choisies aléatoirement.
3. *RAN10* : 10% des sources sont observées lors du suivi. Ces sources étant choisies aléatoirement.

La figure 7.3.5 illustre les contraintes attendues pour les différents suivis considérés.

On constate qu'un suivi partiel introduit une erreur systématique inférieure aux erreurs statistique à un σ . Cependant il sera important d'étudier la fonction de sélection du suivi afin de corriger cette erreur systématique. On peut imaginer évaluer, en fonction du redshift : l'efficacité de détection d'un amas, les erreurs sur la mesure du redshift observé, etc.

7.3.5 Contraintes “robustes”

Les plus grands amas comptent plusieurs centaines de galaxies. Les effets gravitationnels dominent les phénomènes de physique interne (formation d'étoile, AGN, collisions, etc.) ce qui n'est plus vrai pour les amas de basse masse. Pour les amas massifs, la loi d'échelle entre leur masse et leur température présente peu de dispersion. On s'attend donc à ce que les amas massifs soient mieux décrits de manière statistique par les modèles de physique interne utilisés et que les contraintes calculées en se limitant à ces amas soient plus robustes.

La figure 7.3.6 illustre les résultats attendus d'un sondage dont on ne considérerait que les sources les plus brillantes (les amas les plus massifs) dont le flux est supérieur à un flux limite.

Le flux limite doit être choisi avec précaution de façon à être dans le régime où l'efficacité de la chaîne de détection des sources est de 100%. En pratique nous

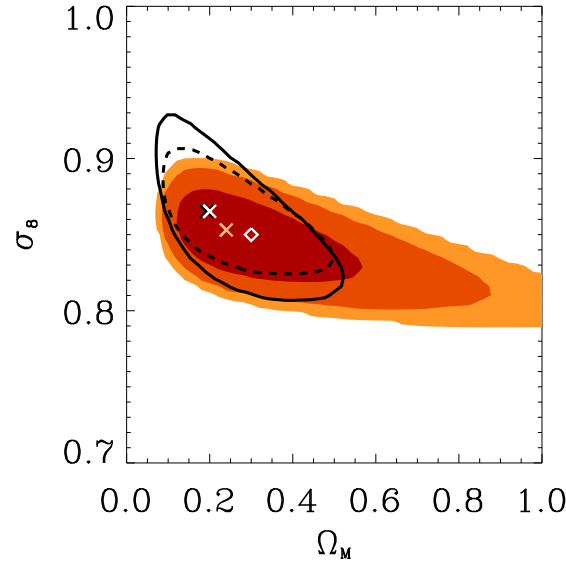


FIG. 7.3.5 – Contraintes attendues avec les suivis partiels *RAN10* et *RAN50*. Les contours grisés (oranges) donnent les niveaux de confiance de référence, en tirêt, l'erreur à un σ pour le suivi *RAN50* et en trait continu l'erreur à un σ pour le suivi *RAN10*.

n'avons considéré que les sources avec $Y > Y_{lim} = 7.4 \times 10^{-4} \text{ arcmin}^2$. Il apparaît dans la figure 7.3.6 (gauche) qu'on introduit une erreur à bas redshift. Par ailleurs on constate qu'un tel choix de limite en flux fait perdre énormément d'information et les erreurs statistiques sur l'estimation des paramètres cosmologiques augmentent considérablement. Cette dégradation importante est due, en partie, au fait que les amas les moins brillants sont aussi les plus lointains et donc portent une information riche sur la cosmologie. Ce résultat illustre l'importance de bien comprendre le comportement de la chaîne de détection pour pouvoir prendre en compte les détections de bas flux et exploiter au mieux les prochaines expériences de sondage grand champ SZ.

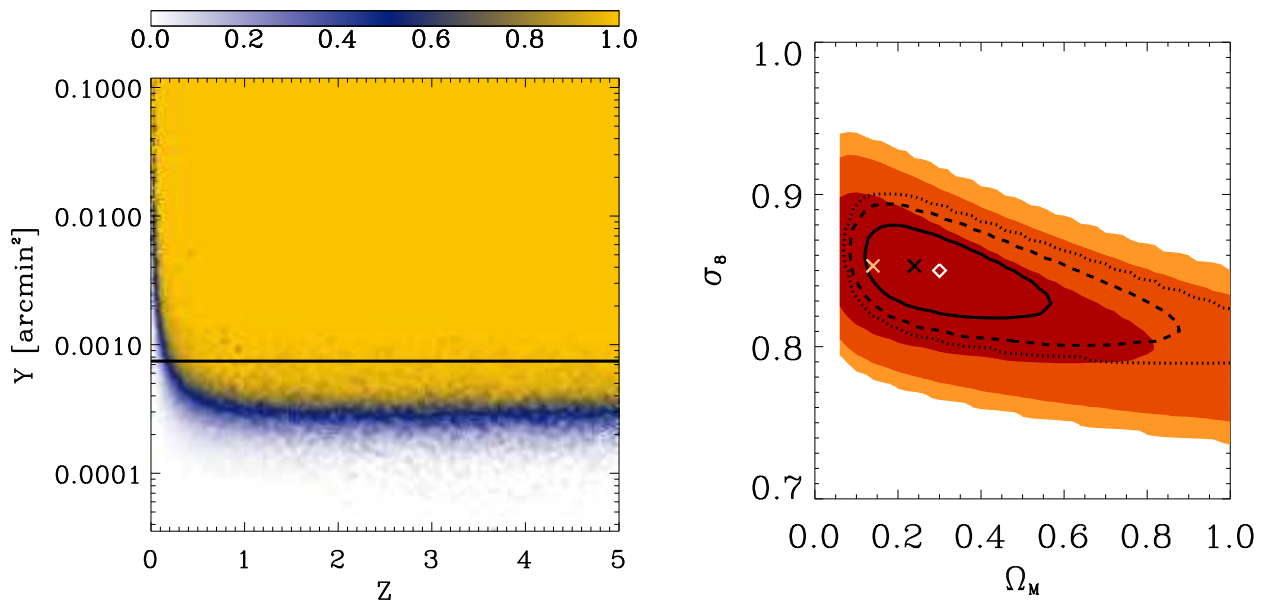


FIG. 7.3.6 – Contraintes avec un suivi *COMPLET* et avec une limite en flux fixée à 7.4×10^{-4} arcmin². (gauche) figure illustrant la coupure en flux superposée à la complétude calculée par Monte-Carlo et utilisée dans le modèle statistique de la chaîne de détection SZ. (droite) Contraintes déduite de ce type d'observation.

7.4 Systématiques liées au modèle statistique

Dans cette section je présente : 1) le biais de reconstruction des paramètres cosmologiques par le modèle statistique, 2) l'influence de la confusion sur ce biais, 3) les erreurs systématiques introduites par une éventuelle mauvaise modélisation de la chaîne de détection SZ.

Les différentes contraintes présentées dans cette section ont été calculées pour l'observation donnée dans la section précédente. Cette observation a été choisie pour être représentative du modèle cosmologique sous-jacent, ceci permet de présenter des contraintes bien centrées autour du modèle de concordance. Par ailleurs, *cette observation sert pour tous les calculs de contraintes*, ainsi il est possible de voir les erreurs *statistiques* et *systématiques* sur les contours présentés. Cependant, dans la section suivante, le biais est calculé grâce à N générations de la chaîne *complète* de simulation-détection.

7.4.1 Estimateur des paramètres cosmologiques

Le premier résultat important est de vérifier que l'estimateur des paramètres cosmologiques que nous avons construit est non biaisé. Cet estimateur est constitué de l'ensemble : *modèle statistique de la chaîne de détection* associé à la *vraisemblance* L (voir l'équation 7.2). La figure 7.4.1 illustre ce résultat : 1) on a généré 100 observations grâce à la simulation présentée au chapitre 4, 2) on utilise le bon modèle statistique (calculé avec une confusion normale) pour estimer les paramètres cosmologiques.

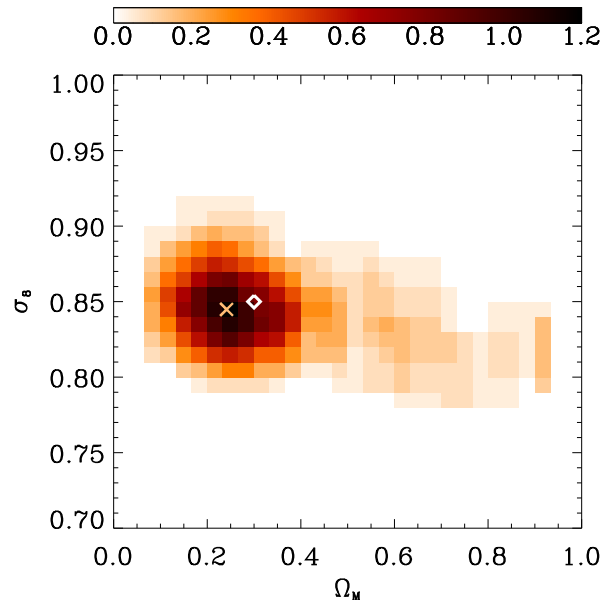


FIG. 7.4.1 – Surface de densité pour N modèles en accord avec N observations issues de simulation complètes d'observation par Olimpo, présentées dans le chapitre 4.

On constate que l'estimateur construit est faiblement biaisé.

L'erreur systématique sur le paramètre σ_8 est inférieure à 1% et inférieure à 2% sur le paramètre Ω_M .

Ayant montré ce biais, on suppose dans toute la suite, que le modèle statistique est valide et introduit un biais négligeable dans l'estimation des paramètres cosmologiques. Cette hypothèse permet alors d'utiliser le modèle statistique pour la génération des observations. On étudie alors l'influence des différentes "composantes" du modèle statistique de la chaîne de détection SZ : complétude, erreurs photométriques, etc. En pratique la différence principale entre les deux approches (simulation complète ou utilisation du modèle statistique) provient des queues dans les distributions des nombres effectifs de détections.

La figure 7.4.2 illustre N générations d'observations tirées grâce au modèle statistique et non plus en simulation complète. Le biais tend vers zéro avec le nombre de générations. L'erreur systématique observée est simplement due au faible nombre de générations.

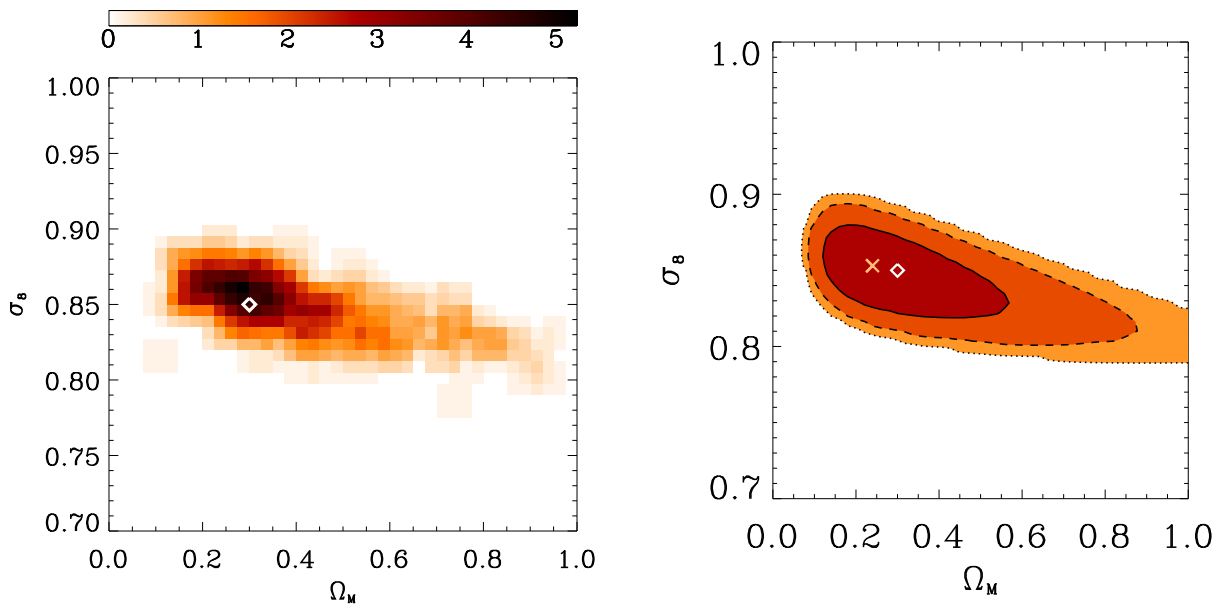


FIG. 7.4.2 – (gauche) Surface de densité pour N modèles en accord avec N observations simulées à partir du modèle statistique de la chaîne de détection SZ. (droite) L'équivalent en termes de contraintes.

7.4.2 Modèle statistique et confusion

Dans ce paragraphe on met en évidence l'influence du bruit de confusion (comme présenté dans le chapitre 6) sur les erreurs faites en termes de contrainte, via, le calcul de deux modèles statistiques de la chaîne de détection : $MS_{1.5}$ et $MS_{0.75}$. Le premier correspond au modèle statistique calculé pour le modèle de concordance avec

une confusion plus importante, le nombre d’amas dans le champ a été artificiellement multiplié par 1.5 et le second avec une confusion plus faible, le nombre d’amas dans le champ a été artificiellement multiplié par 0.75. Ces deux cas ont été présentés en fin de chapitre 6, dans la section 6.3.2.

La confusion est un phénomène susceptible d’introduire une influence importante de la cosmologie sur le modèle statistique. Le fait de la “modéliser” par un changement de la densité de surface des amas est une hypothèse qu’il faudrait vérifier par la suite en étudiant l’influence des modèles cosmologiques sur la taille des détections (elle-même directement reliée à la confusion), phénomène complexe (par exemple, le paramètre σ_8 a une influence sur l’époque de la transition entre les évolutions linéaires et non-linéaires des surdensités et donc sur leur taille à une époque donnée) qui mérite un travail approfondi. Dans un premier temps on se limite à l’étude de l’influence de la densité de surface des amas sur le modèle statistique et sur l’erreur qui en découle dans le calcul des contraintes.

Les résultats sont donnés, comme je l’ai précisé précédemment, *pour la même observation*, dans la figure 7.4.3.

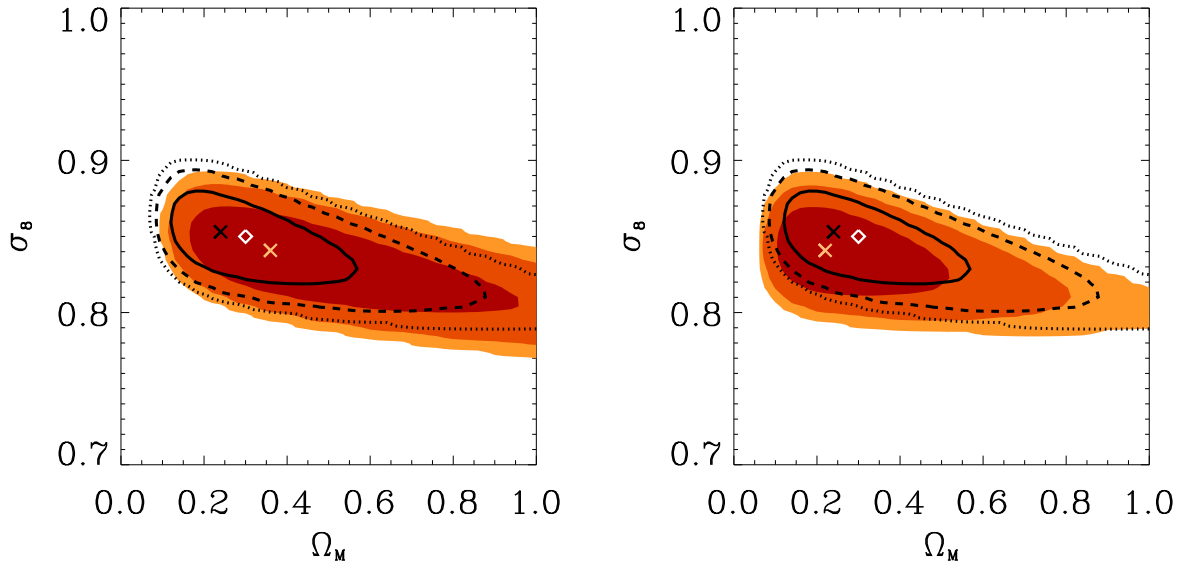


FIG. 7.4.3 – (gauche) Contraintes pour le modèle statistique $MS_{0.75}$. (droite) Contraintes pour le modèle statistique $MS_{1.5}$.

On constate que l’erreur systématique introduite reste dans les erreurs statistiques à un σ .

Les erreurs, sur l’estimation des contraintes, introduites par les différences entre les trois modèles statistiques sont illustrées par la figure 7.4.4. Les contours illustrent les erreurs faites sur les calculs d’erreur à un sigma si on suppose les systématiques nulles. Le calcul est fait grâce aux trois distributions de log-vraisemblance calculées par Monte-Carlo.

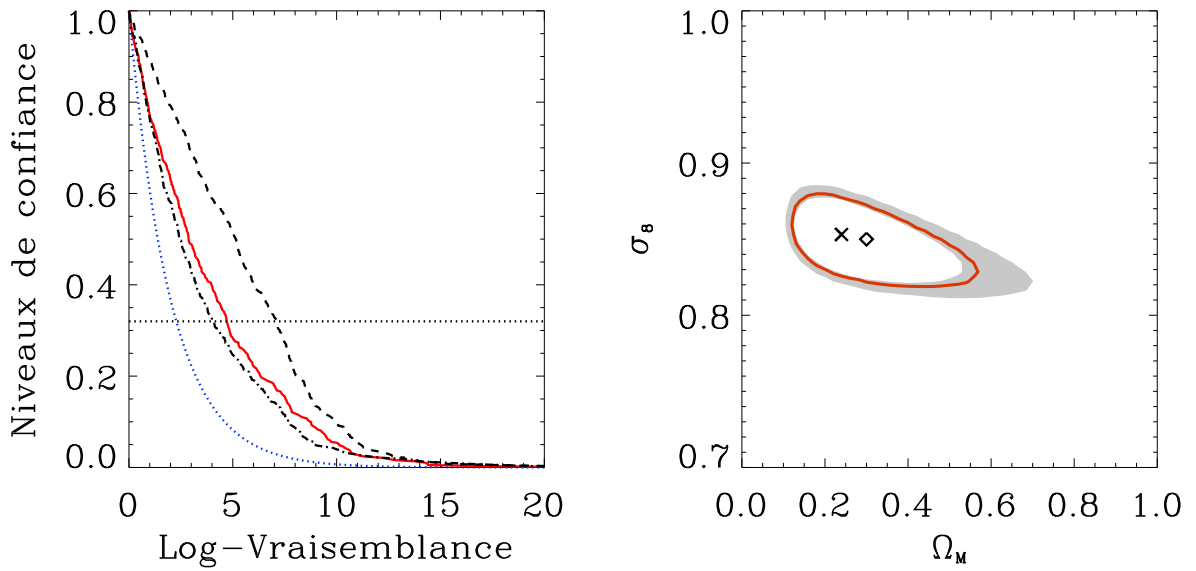


FIG. 7.4.4 – (gauche) Les trois distributions de log-vraisemblance : en trait continu pour le modèle $MS_{1.0}$ en tirets pour le modèle $MS_{0.75}$ et en trait point-tiret pour le modèle $MS_{1.5}$. (droite) Les contours rouges donnent les erreurs statistiques à un sigma pour le modèle $MS_{1.0}$ (la référence), les zones grisées donnent les erreurs sur ces contraintes.

Dans les trois sections suivantes on regarde l'influence d'une "erreur", introduite ici volontairement, dans la modélisation de la chaîne de détection. On introduit des erreurs dans la complétude, les erreurs photométriques et les contaminations.

7.4.3 Complétude

Je modélise l'erreur faite sur la complétude en introduisant, artificiellement, dans le modèle statistique une carte de complétude simplifiée, tous les autres modèles étant identiques par ailleurs (photométrie, contamination, etc). La figure 7.4.5 (gauche) illustre l'erreur faite, volontairement, sur la complétude : l'efficacité de la chaîne de détection est prise à 100% pour tous les flux supérieurs à $2.7 \times 10^{-4} \text{ arcmin}^2$. L'efficacité (calculée par Monte-Carlo), à ce flux limite, est de 50%. Les résultats sur la figure de droite illustrent les contraintes calculées avec ce modèle statistique (toujours pour la même observation). On voit ici l'importance de la complétude dans le calcul des contraintes : l'erreur systématique introduite domine largement les erreurs statistiques sur l'estimation des paramètres.

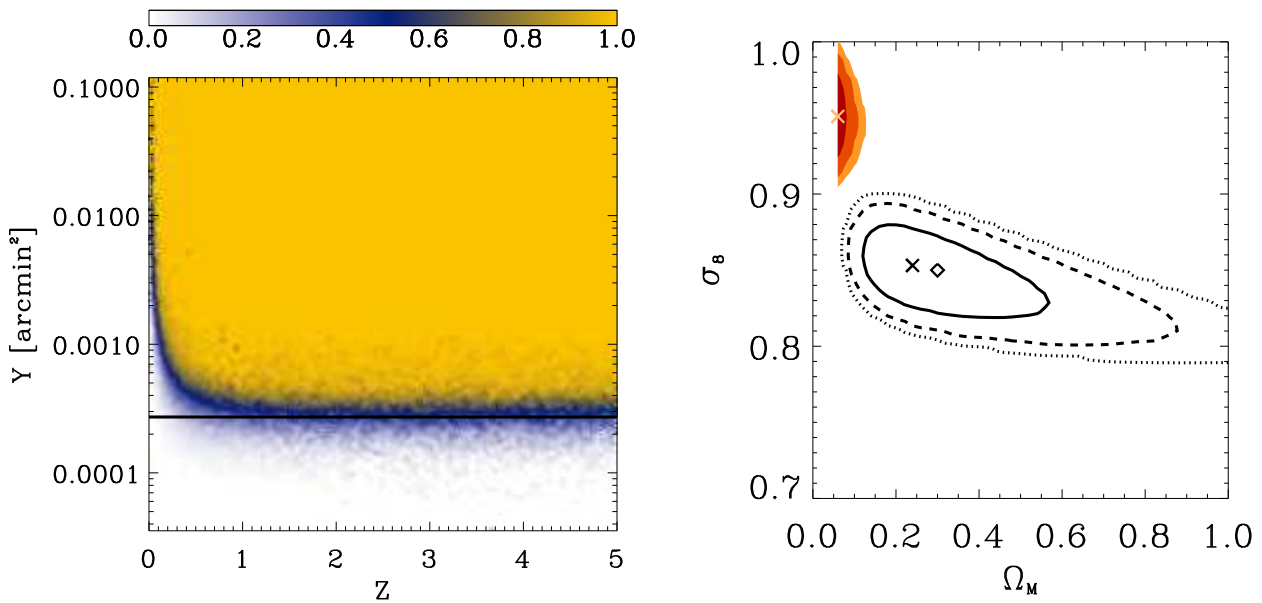


FIG. 7.4.5 – (gauche) Complétude avec une limite en flux fixée à $Y_{lim} = 2.7 \times 10^{-4} \text{ arcmin}^2$ superposée à la complétude calculée par Monte-Carlo. (droite) Contraintes calculées avec ces deux complétudes.

7.4.4 Biais photométrique

La figure 7.4.6 montre l'erreur systématique introduite par le modèle statistique de la chaîne de détection si on ne modélise pas le biais causé par les erreurs photométriques (voir la section 6.2.2). L'impact de cet oubli sur la distribution observée

des détections est très importante : ceci est montré par la différence entre les lignes 2 et 3 de la figure 6.3.1. A bas flux (c'est à dire pour la majorité des détections!) l'erreur est très importante. Ceci explique l'erreur systématique introduite par cet "oubli" dans la modélisation de la chaîne de détection.

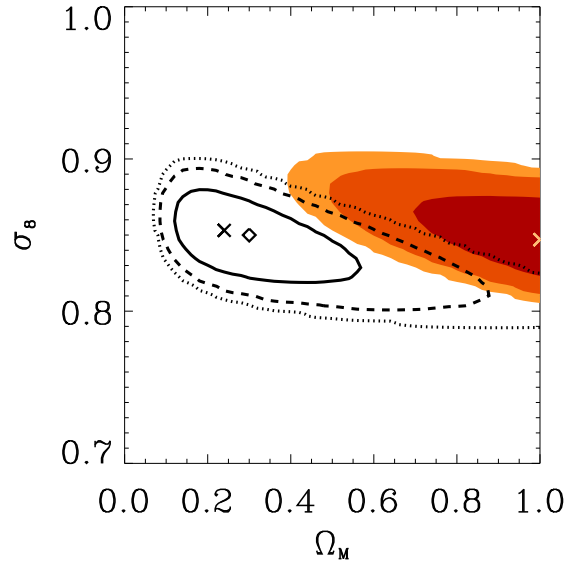


FIG. 7.4.6 – Erreur systématique introduite si on ne tient pas compte du biais photométrique à bas flux dans la fonction de sélection.

7.4.5 Contamination

La figure 7.4.7 présente l'erreur systématique introduite si on ne prend pas en compte la distribution en flux des contaminations dans le modèle statistique. Les observations générées ici sont des observations avec un suivi de type *RAN50*, c'est à dire avec une mesure partielle des redshifts, car si on mesure 100% des redshifts on identifie avec certitude les contaminations et le biais est alors inexistant. Cette hypothèse introduit une erreur systématique relativement faible en σ_8 et Ω_M car les deux paramètres sont bien contraints par la distribution en redshift des amas détectés. L'erreur provient principalement de l'incertitude sur le terme $\frac{dP}{dN^{det}}(N_{exp}^{det}, \vec{P})$ de la vraisemblance mesurant l'accord en termes de nombre de détections (amas et contaminations). Ce terme, mal décrit, introduit une erreur systématique faible comparée aux erreurs statistiques.

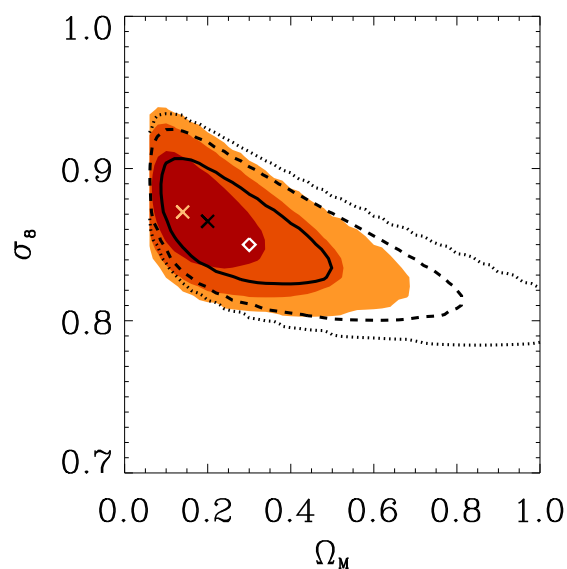


FIG. 7.4.7 – Erreur systématique introduite par un “oubli” de la distribution des contaminations dans le modèle statistique de la chaîne de détection SZ. Les contours de référence en superposition sont ceux calculés pour le suivi de type *RAN50*.

Des trois erreurs présentées ici, la plus influente sur les calculs de contrainte est celle sur la complétude, ensuite celle sur le biais photométrique et enfin sur les contaminations.

7.5 Conclusion

Dans ce chapitre j’ai présenté une perspective des résultats de l’estimation, au premier ordre, des contraintes sur le paramètre T_* grâce au vol scientifique attendu pour Olimpo (figure 7.3.2). Les résultats s’appuient sur les contraintes obtenues (des paramètres Ω_M et σ_8) par les mesures du spectre de puissance du CMB de WMAP et sur une prospective des résultats attendus prochainement sur les mesures de *weak-shear* par l’expérience CFHTLS-wide. Afin d’améliorer encore les contraintes sur l’estimation de T_* il semble important de combiner les futures données SZ avec les données issues d’observation d’amas en X.

Nous avons ensuite considéré T_* bien connu et fixé à la valeur du modèle de concordance. Ceci nous a permis d’étudier différentes systématiques sur le calcul des contraintes sur les paramètres cosmologiques σ_8 et Ω_M . Dans un premier temps nous avons mis en évidence l’impact, sur les contraintes, d’un suivi optique partiel. Nous avons alors constaté une erreur systématique faible qu’on peut espérer corriger en incluant la fonction de sélection propre du suivi optique.

Ces résultats ont été calculés grâce à la mise en place du modèle statistique de la chaîne de détection SZ présenté au chapitre 6. Ce modèle statistique se révèle être un estimateur *faiblement biaisé* des paramètres cosmologiques. Il est intéressant de remarquer qu’une coupure en flux dans le catalogue, conduisant à négliger les amas de faible flux, fait perdre beaucoup en termes de contraintes (figure 7.3.6). Ce modèle statistique se révèle donc important pour exploiter, au mieux, les capacités des prochaines expériences de sondage grand champ SZ.

De plus, l’étude des systématiques montre que ce modèle statistique doit être connu avec une grande précision afin d’estimer correctement les contraintes qui s’imposent aux paramètres cosmologiques. Le point le plus sensible est la complétude, suivi des erreurs photométriques et enfin des contaminations. Les résultats montrent clairement la difficulté *et l’importance* d’une modélisation correcte de la chaîne de détection SZ (figures 7.4.5, 7.4.6 et 7.4.7).

Par ailleurs, l’hypothèse d’indépendance du modèle statistique vis-à-vis des paramètres cosmologiques se trouve, au moins en partie, justifiée par les résultats montrés dans la section 7.4.2. Les erreurs sur les contraintes calculées avec cette hypothèse sont faibles. Nous avons testé seulement le premier ordre de l’influence de la cosmologie, c’est à dire la confusion due à la densité de surface des amas. Il conviendrait pour la suite d’effectuer le même travail en considérant l’influence directe de la cosmologie, en trouvant des paramètres *observables* permettant de paramétriser les modèles statistiques de la chaîne de détection SZ correctement : on peut imaginer que le nombre total de détections et, par exemple, le coefficient de corrélation linéaire de la distribution en flux et redshift (“décrivant” la forme globale de la distribution) soient deux paramètres pertinents.

Conclusion

L'objectif de ce travail de thèse était de préparer l'analyse des données des prochaines observations SZ grand champ. Pour cela, j'ai mis en place une chaîne complète d'outils originaux permettant d'effectuer cette étude. Ces outils ont permis de mettre en évidence des points critiques et essentiels des effets de sélection à prendre en compte dans les futures analyses. Je me suis intéressé au cas des expériences multi-fréquences dotées de grandes caméras de bolomètres. J'ai appliqué, dans ma thèse, ces outils à une perspective pour l'expérience Olimpo.

Simulation de ciel millimétrique

Le premier outil mis en oeuvre est un code de simulation rapide de cartes de ciel observé dans le domaine millimétrique. Il regroupe un code initial de simulation de catalogue d'amas SZ (écrit par A. Réfregier, Dapnia/SAp), la projection de ce catalogue pour en faire une carte d'amas, la simulation de phénomènes astrophysiques contaminants, et la simulation, au premier ordre, d'effets instrumentaux d'observation (ces deux derniers points ont été mis en place par D. Yvon). Le premier point original de ce code concerne la projection du catalogue d'amas SZ sur le ciel. J'ai introduit un profil d'amas, projeté sur la carte, elliptique dont la distribution du paramètre d'ellipticité est déduite d'observations en X d'une cinquantaine d'amas.

La simulation des galaxies infrarouges est originale et permet de mettre en évidence la difficulté de traitement de cette source particulière de contamination. Dans un premier temps, D. Yvon a mis en oeuvre, le fait observé, que chacune des galaxies possède un spectre électromagnétique d'émission propre. Dans un second temps, en se basant sur une constatation observationnelle, il a développé un code permettant de corrélérer la position d'une fraction de ces sources contaminantes avec les positions des amas.

Bien que la simulation inclut ces trois effets astrophysiques originaux je me suis concentré, dans mon travail de thèse, à étudier l'impact des spectres d'indice variable des galaxies infrarouges sur le traitement des données. L'impact des ellipticités sur les erreurs photométriques n'a pas été abordé explicitement, bien que le modèle statistique des effets de sélection en tienne compte.

Avec l'arrivée des données des prochaines expériences, il sera nécessaire de construire une simulation plus réaliste encore, tant au niveau des phénomènes astrophysiques que des effets instrumentaux d'observation. Ayant comme objectif de construire une chaîne complète d'analyse, nous ne sommes pas allés plus loin dans les simulations, ce qui aurait nécessité la connaissance approfondie d'un instrument, notamment pour le passage d'une simulation directe de cartes à la simulation de données ordonnées

en temps.

Méthode de détection d'amas SZ

Le second outil mis en oeuvre est une chaîne d'extraction d'amas SZ à partir des cartes observées. Cet outil doit être capable d'extraire un maximum d'amas, tout en évitant les contaminations, afin d'être en mesure d'extraire un catalogue de détection suffisamment "pur" pour pouvoir espérer contraindre efficacement les paramètres cosmologiques. Par ailleurs, il doit être en mesure d'estimer le flux intégré des amas avec un minimum d'erreurs. Enfin, notre objectif était de construire une chaîne traitant les données "en aveugle", sans *a priori* sur le signal recherché.

Afin de répondre à ces critères j'ai caractérisé, l'un après l'autre, trois algorithmes à l'état-de-l'art du traitement d'image, travaillant en aveugle. La chaîne ainsi constituée, est une approche originale et performante de l'extraction des amas SZ et fait l'objet d'une publication soumise à la revue A&A (Pires et al., 2005). Elle devrait permettre une exploitation optimale des prochaines données et fournir des catalogues comptant plusieurs centaines d'amas dans le cas d'Olimpo.

Outils statistiques et modèle d'observation

Une fois le catalogue de sources extrait et en se plaçant dans l'optique d'analyser les données, j'ai mis en place deux éléments complémentaires rendant cohérente la chaîne d'étude. Ces éléments consistent, tout d'abord, en deux outils d'analyse statistique des catalogues de sources, ensuite, en un modèle statistique des effets de sélection de la chaîne complète de détection d'amas. Ce travail fait l'objet d'une publication soumise à la revue A&A (Juin et al., 2005).

Les outils d'analyse statistique permettent, dans un premier temps de trouver une classe de modèles cosmologiques statistiquement compatible avec l'observation (il est, par exemple, possible de rejeter la fonction de masse proposée par Press & Schechter lorsque les données sont simulées avec celle de Sheth & Tormen) et dans un second temps d'estimer les paramètres cosmologiques. Ces deux outils sont complémentaires et ont été conçus dans l'optique de traiter des observations comptant peu d'amas, ce qui sera le cas dans les premiers mois de prise de données d'APEX ou de Planck, ou pour le vol de test d'Olimpo.

Grâce à ces outils, j'illustre, en particulier, la nécessité de combiner les données SZ, X et de *weak-shear*, en plus des observations du CMB et des supernovae, pour comprendre avec précision la physique interne des amas et pouvoir alors utiliser ceux-ci comme des sondes cosmologiques. Les données simulées d'Olimpo, combinées avec celle de WMAP et du *weak-shear* attendu par l'expérience CFHTLS-*wide*, ne suffisent cependant pas à obtenir des contraintes suffisantes sur les paramètres des modèles de physique interne des amas. Je souligne l'importance d'observations multi-longueurs d'onde d'un champ commun, notamment pour contraindre la paramètre T_* qui quantifie l'efficacité de chauffage du gaz intra-amas.

Par ailleurs, je montre l'impact d'un suivi photométrique *partiel* des sources détectées. Il apparaît qu'un tel suivi n'introduit pas d'erreur systématique dans l'estimation des paramètres cosmologiques.

Ce travail de calcul de contraintes n'est envisageable que par une bonne compréhension des effets de sélection des observations SZ. J'ai proposé dans ma thèse un modèle statistique complet de ces effets. C'est un estimateur "rapide" et faiblement biaisé des paramètres cosmologiques. L'erreur systématique, introduite par l'utilisation de ce modèle, sur les paramètres Ω_M et σ_8 est de l'ordre du pourcent. D'une part, ce modèle permet de s'affranchir d'une étude complète en Monte-Carlo lourde à mettre en place mais surtout il permet une étude détaillée des effets de sélection complexes, que ce soit algorithmiques, instrumentaux ou astrophysiques, liés aux observations SZ grand champ. Par exemple, grâce à ce modèle, j'ai mis en avant, entre autre, la statistique non-poissonienne du nombre effectif de détections attendues pour une observation et son impact sur les calculs de contraintes qui se révèle être un effet dominant.

Contexte scientifique et instrumental

Je rappelle ici l'importance de mon travail dans le contexte actuel des observations SZ. Dans les prochaines années, de 2006 à 2008, trois expériences grand-champ à caméra bolométrique vont entrer en fonction : Olimpo, APEX-SZ et Planck. Les outils que j'ai présenté permettent d'anticiper sur l'analyse des données de ces expériences. Je souligne, en effet, l'importance des différents effets de sélection dans l'estimation des paramètres des modèles cosmologiques.

Par ailleurs, les expériences interférométriques vont de la même manière pouvoir effectuer des observations SZ en grand-champ (par exemple AMI ou SZA). Bien que ces expériences n'aient pas du tout les mêmes caractéristiques que les expériences bolométriques, la démarche que j'ai appliqué pour l'étude des effets de sélection est la même et elle est nécessaire pour extraire l'information scientifique des données en terme de cosmologie. Elle devra être adaptée à ce type d'instrument : le chemin est tracé.

Je tiens à rappeler ici l'importance d'observations communes d'amas à différentes longueurs d'onde. Les observations SZ et X sont complémentaires et nécessaires à une bonne compréhension de la physique du gaz des amas. Par ailleurs les observations en optique, par la mesure de cisaillement gravitationnel, vont permettre de limiter grandement les erreurs systématiques sur l'estimation de la masse des amas et vont apporter une lumière sans précédent sur la compréhension de leur formation et de leur structure interne.

Aujourd'hui on peut envisager sérieusement ce type d'analyse multi-longueurs d'onde sur de grands échantillons statistiques grâce aux instruments en fonctionnement ou en passe de le devenir : XMM et Chandra pour les observations en X, CFHTLS-Mégacam pour les mesures de cisaillement gravitationnel et avec les instruments SZ de seconde génération, Olimpo, Planck, APEX, ACT, AMI, SZA, Bolocam, AMiBA, etc.

L'expérience APEX-SZ

Je tiens à montrer ici l'application que j'ai faite des outils développés durant ma thèse à l'analyse des données de l'expérience APEX. J'ai considéré les caractéristiques d'APEX, équipé de la caméra ASZCA, données dans la figure 3.3.2. Il est bien

évident que la simulation de l'instrument n'est pas réaliste. Elle reprend les différents éléments présentés dans le chapitre 4. La figure 7.5.1 montre la carte d'amas SZ reconstruite après application de la méthode de détection développée pour Olimpo durant ma thèse.

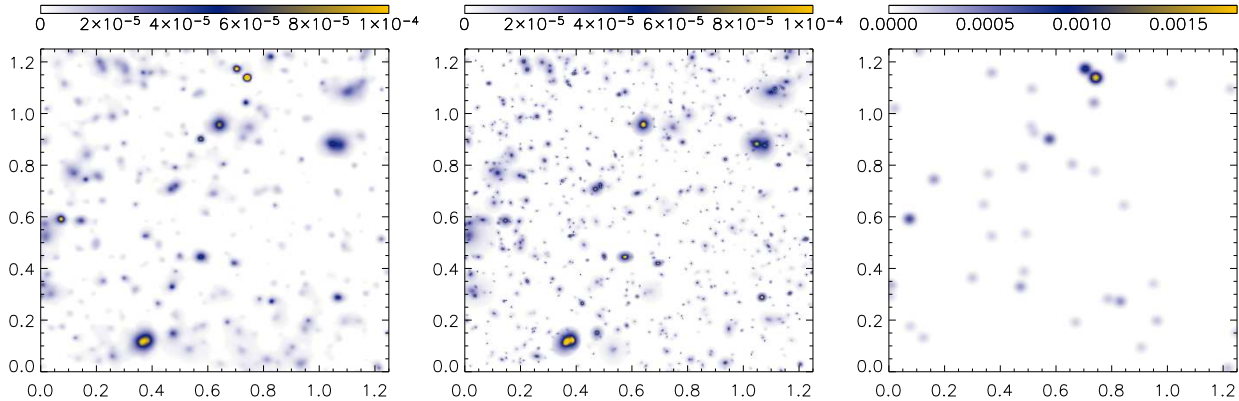


FIG. 7.5.1 – Carte d'amas SZ attendue pour l'expérience APEX-SZ obtenue avec la chaîne de détection décrite dans Pires et al. (2005). La figure de gauche montre la carte SZ reconstruite. La carte au centre montre la carte des amas SZ simulés et à droite la carte des galaxies infrarouges simulée.

La carte SZ reconstruite révèle la présence de sources brillantes qu'on peut directement identifier comme étant des galaxies infrarouges, aux coordonnées (0.7,1.15) par exemple. Celles-ci n'ont pas été isolées de l'effet SZ par l'algorithme ICA travaillant dans les deux canaux d'observation d'ASZCA : 150 et 217 GHz. A la différence de l'expérience Olimpo, la chaîne de traitement des données ne permet pas d'éliminer les sources infrarouges brillantes. La figure 7.5.2 montre les distributions en flux SZ et redshift des sources détectées dans la carte.

Dans le cas d'APEX les sources infrarouges contaminent le catalogue jusqu'au flux $Y \simeq 2 \times 10^{-4}$ arcmin². La plus grande partie des amas ne seront pas identifiables. Dans le cas d'Olimpo, le catalogue atteint une pureté de 90% et les quelques contaminations sont à bas flux. Il apparaît, dans le cas d'APEX, d'autant plus important de séparer les galaxies infrarouges des amas car ceux-ci sont des amas lointains, $z > 1$, riches en information sur l'histoire de l'univers. La figure de droite illustre ce résultat. On constate sur cette figure un effet de sélection intéressant : la résolution angulaire d'APEX permet de sonder les amas sur une plus grande gamme de redshift que pour l'expérience Olimpo.

On constate, à la vue de ces résultats, que la chaîne de détection développée pour Olimpo n'est pas adaptée au traitement des données d'APEX. On peut supposer qu'un pré-traitement des données soit nécessaire. Par ailleurs, on peut imaginer utiliser un des atouts d'APEX, sa résolution angulaire, pour être en mesure *a posteriori* de séparer les deux populations (sources ponctuelles, les galaxies, et étendues, les amas) sur des critères de forme. Il est légitime de se poser la question des performances, pour Olimpo, de cette même chaîne de détection pour isoler les amas des sources ponctuelles radios ? Ces sources radios, encore mal comprises

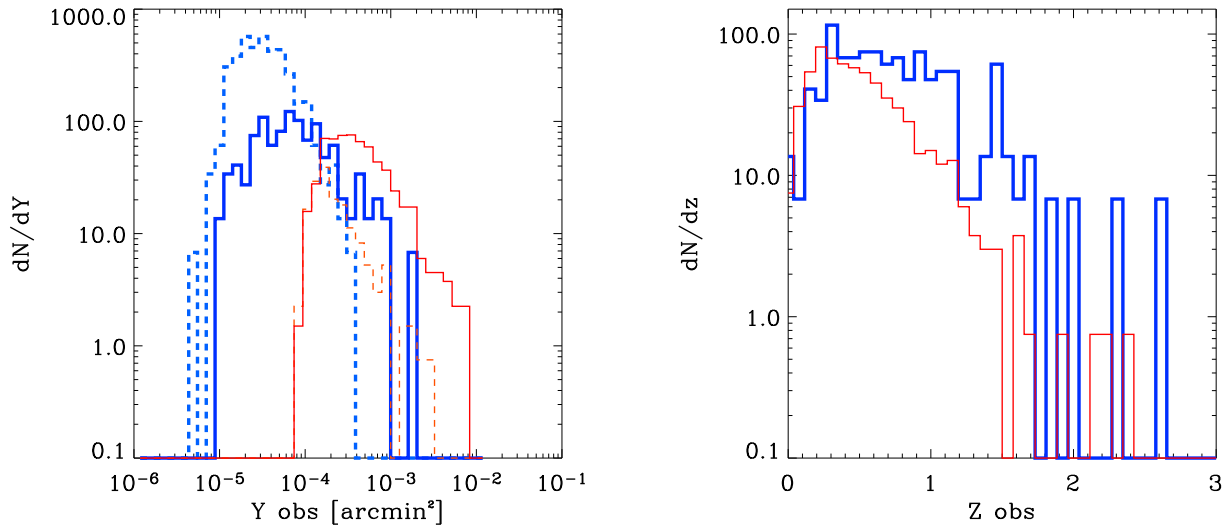


FIG. 7.5.2 – Les histogrammes épais (bleus) sont les distributions des amas (trait continu) et des contaminations (tirêts) pour l'expérience ASZCA tandis que les courbes fines (rouges) sont les mêmes distributions pour l'expérience Olimpo. La figure de droite montre la distribution en redshift des amas détectés pour les deux expériences.

aux fréquences millimétriques, présentent, tout comme les sources infrarouges, un spectre d'émission individuel et constituent *a priori* une source importante de contamination pour les catalogue d'amas SZ.

Ce travail exploratoire met en évidence l'importance du travail de compréhension des effets de sélection pour exploiter au mieux les prochains instruments d'observation de grands champs d'amas SZ. Il montre aussi le travail important fourni, et à fournir, pour la mise au point d'algorithmes de traitement du signal permettant d'extraire un maximum d'information des données provenant des instruments de mesure. Ces derniers, eux-même de plus en plus performants, ouvrent les portes d'une meilleure compréhension de l'univers qui nous entoure, nous réservant, sans aucun doute, encore bien des surprises : 95% du contenu de l'univers restant encore inconnu aujourd'hui.

Table des figures

1.3.1 Les anisotropies du CMB par WMAP	30
1.3.2 Les grandes structures vues par 2dF	30
2.1.1 Exemple d'amas de galaxies en X, SZ et optique	35
2.2.1 Effondrement sphérique : effondrement d'une sphère de masse constante.	40
2.2.2 Evolution de la densité dans le modèle d'effondrement sphérique	41
2.2.3 Distribution de l'ellipticité apparente des amas de galaxies	47
2.4.1 Dépendance spectrale de l'effet SZ thermique relativiste et non-relativiste.	55
3.1.1 Sondages SZ d'actualité.	61
3.1.2 Sondages d'amas grand-champ dans différentes longueurs d'onde.	62
3.2.1 Schéma et photo d'Olimpo	64
3.2.2 Schéma du cryostat d'Olimpo	65
3.3.1 Photo et schéma des expériences APEX et Planck	67
3.3.2 Caractéristiques comparatives des expériences Olimpo, APEX et Planck	68
4.1.1 Les quatre phénomènes physiques principaux dans le domaine sub-millimétrique	77
4.2.1 Exemple de cartes simulées d'amas de galaxies observés par effet SZ	79
4.2.2 Exemple de génération d'un catalogue d'amas	82
4.3.1 Exemple de simulation d'une carte de CMB	87
4.3.2 Simulation d'une carte de galaxies infrarouges	88
4.3.3 Exemple de carte simulée d'émission thermique de la poussière galactique	89
4.4.1 Schéma des spectres d'émissivité des composantes physiques simulées	90
4.5.1 Cartes simulées du ciel aux quatre fréquences : 143, 217, 385 et 600 GHz	94
5.1.1 Les amas, simulés, et vus aux quatre fréquences d'Olimpo : détail des cartes.	101
5.2.1 Cartes reconstruites par JADE et signaux physiques	107
5.2.2 Cartes d'erreur entre la carte SZ reconstruite et la carte simulée avec et sans décomposition dans la base des ondelettes de Mallat	110
5.2.3 Distorsion D^l du signal SZ reconstruit en fonction de l'échelle sur la carte	111

5.3.1 Cartes d'effet SZ simulée et reconstruites, présentées pour différents filtrages	119
5.3.2 Résultats des filtrages gaussien et ME-FDR	121
5.3.3 Comparaison du filtrage ME-FDR au filtrage gaussien	122
5.4.1 Cartes SZ aux différentes étapes de la chaîne de simulation-détection	124
5.4.2 Histogrammes des distances entre sources détectées et amas simulés	126
5.4.3 Pureté et complétude des catalogue reconstruits grâce aux différents filtrages	128
5.5.1 Schéma de la chaîne de détection d'amas SZ	133
6.2.1 Distribution des amas simulés et détectés en fonction de leur redshift, flux et masse simulés	141
6.2.2 Complétude d'un sondage type Olimpo	142
6.2.3 Complétude d'un sondage de type Olimpo : Monte-Carlo et modèle statistique	143
6.2.4 Reconstruction de la taille des amas détectés	145
6.2.5 Estimation du paramètre de Compton intégré Y_{obs} pour les amas détectés	146
6.2.6 Calcul du modèle statistique d'erreur photométrique : $Y_{obs}(Y_{th})$	147
6.2.7 Modèle statistique des erreurs photométriques sur l'estimation de Y_{obs}	148
6.2.8 Distribution des contaminations en fonction de leur flux estimé Y_{obs}	150
6.2.9 Distributions du nombre de sources détectées calculée par Monte-Carlo	152
6.3.1 Validation du modèle statistique : complétude et erreurs photométriques	155
6.3.2 Validation du modèle statistique : distributions en flux et contaminations	156
6.3.3 Complétude pour les différents niveaux de confusion	158
6.3.4 Confusion et erreurs photométriques	159
6.3.5 Confusion et contaminations	160
6.3.6 Confusion et nombre d'amas	161
6.3.7 Confusion et nombre de contaminations	162
7.1.1 Schéma illustrant le travail d'analyse statistique d'une observation	166
7.2.1 Comparaison de la distribution théorique des amas en flux et redshift et du catalogue observé.	168
7.2.2 Tests d'hypothèse par Monte-Carlo	171
7.2.3 Tests d'hypothèse : sélectionner la fonction de masse	172
7.2.4 Etapes de calcul des contraintes avec une distribution de χ^2	174
7.2.5 Comparaison des distributions des log-vraisemblance théorique et calculée par Monte-Carlo	175
7.2.6 Contraintes sur Ω_M et σ_8 : influence du nombre et de la distribution en flux des amas	176
7.3.1 Observation typique d'un vol scientifique d'Olimpo.	178
7.3.2 Contraintes sur les paramètres σ_8 et T_\star	180
7.3.3 Contraintes sur les paramètres Ω_M et σ_8 attendues pour le vol scientifique d'Olimpo	182
7.3.4 Dégradation des contraintes due à la variance des comptages	183
7.3.5 Contraintes attendues avec les suivis partiels $RAN10$ et $RAN50$	185

7.3.6 Contraintes obtenues avec une simple limite en flux	186
7.4.1 Un estimateur faiblement biaisé.	187
7.4.2 Le modèle statistique supposé non-biaisé.	188
7.4.3 Contraintes pour les deux modèles statistique à faible et forte confusion	189
7.4.4 Erreur sur les erreurs statistiques	190
7.4.5 Erreur systématique introduite par un modèle de complétude avec une limite en flux mal placée.	191
7.4.6 Erreur systématique due au biais photométrique	192
7.4.7 Erreur systématique introduite par une mauvaise modélisation des contaminations	193
7.5.1 Carte d'amas SZ attendue pour l'expérience APEX-SZ grâce à la chaîne de détection décrite dans Pires et al. (2005)	198
7.5.2 Distribution des sources détectées pour l'expérience APEX-SZ	199

Liste des tableaux

1.1	Paramètres de l'équation d'état du contenu de l'univers	12
1.2	Les paramètres cosmologiques aujourd'hui	15
2.1	Grandeurs caractéristiques des amas de galaxies	38
3.1	Caractéristiques d'Olimpo	66
4.1	Paramètres cosmologiques utilisés pour les exemples du chapitre . . .	80
5.1	Paramètres SExtractor utilisés pour l'extraction des amas	123
6.1	Modèle cosmologique de concordance	137
7.1	Paramètres Ω_M et σ_8 : erreurs statistiques à un σ	179

Bibliographie

- G. O. Abell. The Distribution of Rich Clusters of Galaxies. *ApJS*, 3 :211–+, May 1958.
- N. Aghanim, K. M. Górski, and J.-L. Puget. How accurately can the SZ effect measure peculiar cluster velocities and bulk flows? *A&A*, 374 :1–12, July 2001. doi : 10.1051/0004-6361 :20010659.
- S. W. Allen, R. W. Schmidt, and A. C. Fabian. The X-ray virial relations for relaxed lensing clusters observed with Chandra. *MNRAS*, 328 :L37–L41, December 2001. doi : 10.1046/j.1365-8711.2001.05079.x.
- M. Arnaud. Structural and Scaling Properties of the X-Ray Cluster Population and their Cosmological Implications. In *Astronomical Society of the Pacific Conference Series*, pages 85–+, 2003.
- M. Arnaud, N. Aghanim, E. Belsole, R. Gastaud, S. Majerowicz, D. M. Neumann, and J. L. Sauvageot. First results on clusters observed by XMM/EPIC. *Highlights in Astronomy*, 12 :507–509, 2002.
- M. Arnaud, G. W. Pratt, and E. Pointecouteau. Structural and scaling properties of galaxy clusters : probing the physics of structure formation. *Memorie della Societa Astronomica Italiana*, 75 :529–+, 2004.
- S. Arnouts, L. Moscardini, E. Vanzella, S. Colombi, S. Cristiani, A. Fontana, E. Giallongo, S. Matarrese, and P. Saracco. Measuring the redshift evolution of clustering : the Hubble Deep Field South. *MNRAS*, 329 :355–366, January 2002. doi : 10.1046/j.1365-8711.2002.04988.x.
- N. A. Bahcall, J. P. Ostriker, S. Perlmutter, and P. J. Steinhardt. The Cosmic Triangle : Revealing the State of the Universe. *Science*, 284 :1481–+, May 1999.
- D. Barbosa, J. G. Bartlett, A. Blanchard, and J. Oukbir. The Sunyaev-Zel’dovich effect and the value of $\{\Omega_{\text{M}}\}_0$. *A&A*, 314 :13–17, October 1996.
- J. M. Bardeen, J. R. Bond, N. Kaiser, and A. S. Szalay. The statistics of peaks of Gaussian random fields. *ApJ*, 304 :15–61, May 1986. doi : 10.1086/164143.

- J. G. Bartlett. The standard cosmological model and CMB anisotropies. *New Astronomy Review*, 43 :83–109, July 1999.
- R. A. Battye and J. Weller. Constraining cosmological parameters using Sunyaev-Zel'dovich cluster surveys. *Phys. Rev. D*, 68(8) :083506–+, October 2003.
- Y. Benjamini and Y Hochberg. Controlling the false discovery rate - a practical and powerful approach to multiple testing. *J. R. Stat. Soc. B*, 57 :289, 1995.
- E. Bertin and S. Arnouts. SExtractor : Software for source extraction. *A&AS*, 117 : 393–404, June 1996.
- J. Binney and S. Tremaine. *Galactic dynamics*. Princeton, NJ, Princeton University Press, 1987, 747 p., 1987a.
- J. Binney and S. Tremaine. Book Review : Galactic dynamics. / Princeton U Press, 1988. *Nature*, 326 :219–+, March 1987b.
- M. Birkinshaw. The Sunyaev-Zel'dovich effect. *Phys. Rep.*, 310 :97–195, March 1999.
- A. Blanchard and J. G. Bartlett. What does cluster redshift evolution reveal? *A&A*, 332 :L49–L52, April 1998.
- C. Borys, S. Chapman, M. Halpern, and D. Scott. The Hubble Deep Field North SCUBA Super-map - I. Submillimetre maps, sources and number counts. *MNRAS*, 344 :385–398, September 2003. doi : 10.1046/j.1365-8711.2003.06818.x.
- F. R. Bouchet and R. Gispert. Foregrounds and CMB experiments I. Semi-analytical estimates of contamination. *New Astronomy*, 4 :443–479, September 1999. doi : 10.1016/S1384-1076(99)00027-5.
- F. R. Bouchet, R. Gispert, N. Aghanim, J. R. Bond, A. de Luca, E. Hivon, and B. Maffei. Simulations of the Microwave Sky and of Its "Observations". *Space Science Reviews*, 74 :37–43, November 1995.
- E. T. Byram, T. A. Chubb, and H. Friedman. Cosmic X-Ray Sources-Galactic and Extragalactic. *AJ*, 71 :379–+, August 1966.
- Jean-François Cardoso. High-order contrasts for independent component analysis. *Neural Computation*, 11(1) :157–192, January 1999.
- Jean-François Cardoso. Blind signal separation : statistical principles. *Proceedings of the IEEE. Special issue on blind identification and estimation*, 9(10) :2009–2025, oct 1998.
- R. G. Carlberg, H. K. C. Yee, E. Ellingson, R. Abraham, P. Gravel, S. Morris, and C. J. Pritchet. Galaxy Cluster Virial Masses and Omega. *ApJ*, 462 :32–+, May 1996. doi : 10.1086/177125.

- J. E. Carlstrom, M. Joy, L. Grego, G. Holder, W. L. Holzapfel, S. LaRoque, J. J. Mohr, and E. D. Reese. The Sunyaev-Zel'dovich Effect : Results and Future Prospects. *ArXiv Astrophysics e-prints*, March 2001.
- S. Colafrancesco. SZ effect from Dark Matter annihilation. *A&A*, 422 :L23–L27, July 2004. doi : 10.1051/0004-6361 :20040175.
- S. Colafrancesco. The SZ effect from cluster cavities. *A&A*, 435 :L9–L12, May 2005. doi : 10.1051/0004-6361 :200500100.
- S. Colafrancesco, P. Marchegiani, and E. Palladino. The non-thermal Sunyaev-Zel'dovich effect in clusters of galaxies. *A&A*, 397 :27–52, January 2003. doi : 10.1051/0004-6361 :20021199.
- A. Cooray, W. Hu, and J. Miralda-Escudé. Weak Lensing by Large-Scale Structure : A Dark Matter Halo Approach. *ApJ*, 535 :L9–L12, May 2000. doi : 10.1086/312696.
- A. R. Cooray. Galaxy clusters : oblate or prolate? *MNRAS*, 313 :783–788, April 2000.
- A. C. da Silva, S. T. Kay, A. R. Liddle, and P. A. Thomas. Hydrodynamical simulations of the Sunyaev-Zel'dovich effect : cluster scaling relations and X-ray properties. *MNRAS*, 348 :1401–1408, March 2004. doi : 10.1111/j.1365-2966.2004.07463.x.
- P. de Bernardis, P. A. R. Ade, J. J. Bock, J. R. Bond, J. Borrill, A. Boscaleri, K. Coble, B. P. Crill, G. De Gasperis, P. C. Farese, P. G. Ferreira, K. Ganga, M. Giacometti, E. Hivon, V. V. Hristov, A. Iacoangeli, A. H. Jaffe, A. E. Lange, L. Martinis, S. Masi, P. V. Mason, P. D. Mauskopf, A. Melchiorri, L. Miglio, T. Montroy, C. B. Netterfield, E. Pascale, F. Piacentini, D. Pogosyan, S. Prunet, S. Rao, G. Romeo, J. E. Ruhl, F. Scaramuzzi, D. Sforna, and N. Vittorio. A flat Universe from high-resolution maps of the cosmic microwave background radiation. *Nature*, 404 :955–959, April 2000.
- S. De Grandi and S. Molendi. Temperature Profiles of Nearby Clusters of Galaxies. *ApJ*, 567 :163–177, March 2002. doi : 10.1086/338378.
- Jacques Delabrouille, Jean-François Cardoso, and Guillaume Patanchon. Multi-detector multi-component spectral matching and applications for CMB data analysis. *Monthly Notices of the Royal Astronomical Society*, 346(4) :1089–1102, December 2003. to appear, also available as <http://arXiv.org/abs/astro-ph/0211504>.
- A. Diaferio, A. Nusser, N. Yoshida, and R. A. Sunyaev. Superclusters with thermal Sunyaev-Zel'dovich effect surveys. *MNRAS*, 338 :433–442, January 2003. doi : 10.1046/j.1365-8711.2003.06060.x.
- R. H. Dicke, P. J. E. Peebles, P. G. Roll, and D. T. Wilkinson. Cosmic Black-Body Radiation. *ApJ*, 142 :414–419, July 1965. doi : 10.1086/148306.

- J. M. Diego, P. Vielva, E. Martínez-González, J. Silk, and J. L. Sanz. A Bayesian non-parametric method to detect clusters in Planck data. *MNRAS*, 336 :1351–1363, November 2002.
- M. Disney. Book Review : Principles of physical cosmology / Princeton U Press, 1993. *Nature*, 365 :117–+, September 1993.
- H. Dole. L'évolution des galaxies infrarouges des observations cosmologiques avec ISO à une modélisation de l'infrarouge moyen au submillimétrique, 2000.
- M. Doran. CMBEASY : : an Object Oriented Code for the Cosmic Microwave Background. *ArXiv Astrophysics e-prints*, February 2003.
- A. E. Evrard, T. J. MacFarland, H. M. P. Couchman, J. M. Colberg, N. Yoshida, S. D. M. White, A. Jenkins, C. S. Frenk, F. R. Pearce, J. A. Peacock, and P. A. Thomas. Galaxy Clusters in Hubble Volume Simulations : Cosmological Constraints from Sky Survey Populations. *ApJ*, 573 :7–36, July 2002. doi : 10.1086/340551.
- D. P. Finkbeiner and D. J. Schlegel. Interstellar Dust Emission as a CMB Foreground. In *ASP Conf. Ser. 181 : Microwave Foregrounds*, pages 101–+, 1999.
- A. Finoguenov, M. Arnaud, and L. P. David. Temperature and Heavy-Element Abundance Profiles of Cool Clusters of Galaxies from ASCA. *ApJ*, 555 :191–204, July 2001a. doi : 10.1086/321457.
- A. Finoguenov, T. H. Reiprich, and H. Böhringer. Details of the mass-temperature relation for clusters of galaxies. *A&A*, 368 :749–759, March 2001b. doi : 10.1051/0004-6361 :20010080.
- R. Giacconi, S. Murray, H. Gursky, E. Kellogg, E. Schreier, and H. Tananbaum. The UHURU catalog of X-ray sources. *ApJ*, 178 :281–308, December 1972.
- R. Giacconi, S. Murray, H. Gursky, E. Kellogg, E. Schreier, T. Matilsky, D. Koch, and H. Tananbaum. The Third UHURU Catalog of X-Ray Sources. *ApJS*, 27 : 37–+, February 1974.
- L. Grego, J. E. Carlstrom, E. D. Reese, G. P. Holder, W. L. Holzapfel, M. K. Joy, J. J. Mohr, and S. Patel. Galaxy Cluster Gas Mass Fractions from Sunyaev-Zeldovich Effect Measurements : Constraints on Ω_M . *ApJ*, 552 :2–14, May 2001. doi : 10.1086/320443.
- R. Gribonval, L. Benaroya, E. Vincent, and C. Févotte. Performance Measurement in Source Separation. *IRISA publication interne*, 2002.
- K. Griest. Galactic microlensing as a method of detecting massive compact halo objects. *ApJ*, 366 :412–421, January 1991. doi : 10.1086/169575.
- S.F. Gull and J. Skilling. *MEMSYS5 Quantified Maximum Entropy User's Manual*. Royston, England, 1991.

- A. H. Guth. Inflationary universe : A possible solution to the horizon and flatness problems. *Phys. Rev. D*, 23 :347–356, January 1981.
- D. Herranz, J. L. Sanz, R. B. Barreiro, and E. Martínez-González. Scale-adaptive Filters for the Detection/Separation of Compact Sources. *ApJ*, 580 :610–625, November 2002a. doi : 10.1086/342651.
- D. Herranz, J. L. Sanz, M. P. Hobson, R. B. Barreiro, J. M. Diego, E. Martínez-González, and A. N. Lasenby. Filtering techniques for the detection of Sunyaev-Zel’dovich clusters in multifrequency maps. *MNRAS*, 336 :1057–1068, November 2002b. doi : 10.1046/j.1365-8711.2002.05704.x.
- M. P. Hobson and C. McLachlan. A Bayesian approach to discrete object detection in astronomical data sets. *MNRAS*, 338 :765–784, January 2003. doi : 10.1046/j.1365-8711.2003.06094.x.
- A. M. Hopkins, C. J. Miller, A. J. Connolly, C. Genovese, R. C. Nichol, and L. Wasserman. A New Source Detection Algorithm Using the False-Discovery Rate. 123 : 1086–1094, feb 2002.
- D. Huterer and M. White. Weak Lensing as a Calibrator of the Cluster Mass-Temperature Relation. *ApJ*, 578 :L95–L98, October 2002. doi : 10.1086/344699.
- A. Hyvarinen, J. Karhunen, and E. Oja. *Independent Component Analysis*. John Wiley, New York, 2001. URL <http://www.cis.hut.fi/projects/ica/book/>. 481+xxii pages.
- A. Jenkins, C. S. Frenk, S. D. M. White, J. M. Colberg, S. Cole, A. E. Evrard, H. M. P. Couchman, and N. Yoshida. The mass function of dark matter haloes. *MNRAS*, 321 :372–384, February 2001.
- Y. P. Jing and Y. Suto. The Density Profiles of the Dark Matter Halo Are Not Universal. *ApJ*, 529 :L69–L72, February 2000. doi : 10.1086/312463.
- C. Jones and W. Forman. The structure of clusters of galaxies observed with Einstein. *ApJ*, 276 :38–55, January 1984. doi : 10.1086/161591.
- J.-B. Juin, D. Yvon, A. Réfrégier, and C. Yèche. astro-ph/0512378. *submitted to A&A*, 2005.
- G. Jungman, M. Kamionkowski, and K. Griest. hep-ph/9506380. 1995.
- A. Kashlinsky, J. Mather, and S. Odenwald. Fluctuations in the near-IR cosmic infrared background from COBE DIRBE maps. *American Astronomical Society Meeting Abstracts*, 194 :–+, September 1999.
- I. R. King. Density Data and Emission Measure for a Model of the Coma Cluster. *ApJ*, 174 :L123+, June 1972.
- T. Kitayama. Cosmological Implications of Abundances of Clusters of Galaxies, 1997.

- R. Kneissl, M. E. Jones, R. Saunders, V. R. Eke, A. N. Lasenby, K. Grainge, and G. Cotter. Surveying the sky with the Arcminute MicroKelvin Imager : expected constraints on galaxy cluster evolution and cosmology. *MNRAS*, 328 :783–794, December 2001. doi : 10.1046/j.1365-8711.2001.04815.x.
- C. L. Kuo, P. A. R. Ade, J. J. Bock, C. Cantalupo, M. D. Daub, J. Goldstein, W. L. Holzapfel, A. E. Lange, M. Lueker, M. Newcomb, J. B. Peterson, J. Ruhl, M. C. Runyan, and E. Torbet. High-Resolution Observations of the Cosmic Microwave Background Power Spectrum with ACBAR. *ApJ*, 600 :32–51, January 2004. doi : 10.1086/379783.
- E. Kuruoglu, L. Bedini, M. Paratore, E. Salerno, and A. Tonazzini. Source separation in astrophysical maps using independent factor analysis. *Neural Networks*, 16(3-4) :479–491, 2003.
- C. Lacey and S. Cole. Merger rates in hierarchical models of galaxy formation. *MNRAS*, 262 :627–649, June 1993.
- G. Lagache, A. Abergel, F. Boulanger, and J.-L. Puget. The interstellar cold dust observed by COBE. *A&A*, 333 :709–720, May 1998.
- G. Lagache, J. Puget, A. Abergel, F. Desert, H. Dole, F. R. Bouchet, F. Boulanger, P. Ciliegi, D. L. Clements, C. Cesarsky, D. Elbaz, A. Franceschini, R. Gispert, B. Guiderdoni, L. M. Haffner, M. Harwit, R. Laureijs, D. Lemke, A. F. M. Moorwood, S. Oliver, W. T. Reach, R. J. Reynolds, M. Rowan-Robinson, M. Stickel, and S. L. Tufte. The Extragalactic Background and Its Fluctuations in the Far-Infrared Wavelengths. *LNP Vol. 548 : ISO Survey of a Dusty Universe*, 548 : 81–+, 2000.
- J.M. Lamarre, J.L. Puget, F. Bouchet, P.A.R. Ade, and A. Benoit et al. The planck high frequency instrument, a 3rd generation cmb experiment, and a full sky sub-millimeter survey. *New Astronomy Reviews*, 2004. Proceedings of the workshop on "The Cosmic Microwave Background and its Polarization", eds., S. Hanany and R.A. Olive.
- H. Lampeitl, J. Adelman-McCarthy, J. Barentine, A. Becker, W. Boroski, H. Brewington, A. Connolly, F. DeJongh, J. Dembicky, B. Dilday, M. Doi, J. Frieman, J. Gunn, M. Harvanek, S. Hawley, J. Hendry, P. Hoefflich, C. Hogan, J. Holtzman, J. Inkmann, D. Johnston, J. Kaplan, R. Kessler, B. Ketzeback, G. Kilper, A. Kleinman, S. Kleinman, R. G. Kron, S. Krughoff, J. Krzesinski, D. Lamenti, D. Long, J. Marriner, R. McMillan, G. Miknaitis, P. R. Newman, R. Nichol, A. Riess, R. Romani, M. Sako, R. Scranton, S. Snedden, C. Stoughton, M. Subbarao, D. Tucker, L. Wang, N. Yasuda, D. R. Yocum, D. York, and Sloan Digital Sky Survey. First results from the SDSS Supernova Survey. *American Astronomical Society Meeting Abstracts*, 205 :–+, December 2004.
- E. L. Łokas and G. A. Mamon. Dark matter distribution in the Coma cluster from galaxy kinematics : breaking the mass-anisotropy degeneracy. *MNRAS*, 343 : 401–412, August 2003. doi : 10.1046/j.1365-8711.2003.06684.x.

- C. Ma and J. N. Fry. What Does It Take to Stabilize Gravitational Clustering? *ApJ*, 538 :L107–L111, August 2000. doi : 10.1086/312819.
- D. S. Madgwick, E. Hawkins, O. Lahav, S. Maddox, P. Norberg, J. A. Peacock, I. K. Baldry, C. M. Baugh, J. Bland-Hawthorn, T. Bridges, R. Cannon, S. Cole, M. Colless, C. Collins, W. Couch, G. Dalton, R. De Propris, S. P. Driver, G. Efsthliou, R. S. Ellis, C. S. Frenk, K. Glazebrook, C. Jackson, I. Lewis, S. Lumsden, B. A. Peterson, W. Sutherland, and K. Taylor. The 2dF Galaxy Redshift Survey : galaxy clustering per spectral type. *MNRAS*, 344 :847–856, September 2003. doi : 10.1046/j.1365-8711.2003.06861.x.
- D. Maino, A. Farusi, C. Baccigalupi, F. Perrotta, and A. Banday. All-sky astrophysical component separation with fast independent component analysis (fastica). *Monthly Notices of the Royal Astronomical Society*, 334(1) :53–68, 2002.
- K. Maisinger, M. P. Hobson, and A. N. Lasenby. Maximum-entropy image reconstruction using wavelets. *MNRAS*, 347 :339–354, jan 2004.
- P. J. Marshall, M. P. Hobson, S. F. Gull, and S. L. Bridle. Maximum-entropy weak lens reconstruction : improved methods and application to data. *MNRAS*, 335 :1037–1048, oct 2002.
- S. Masi, P. A. R. Ade, R. Artusa, J. J. Bock, A. Boscaleri, B. P. Crill, P. de Bernardis, G. de Troia, P. C. Farese, M. Giacometti, V. V. Hristov, A. Iacoangeli, A. E. Lange, A. T. Lee, L. Martinis, P. V. Mason, P. D. Mauskopf, F. Melchiorri, L. Miglio, T. Montroy, C. B. Netterfield, E. Pascale, F. Piacentini, P. L. Richards, G. Romeo, J. E. Ruhl, and F. Scaramuzzi. BOOMERanG : A Scanning Telescope for 10 Arcminutes Resolution CMB Maps. In *AIP Conf. Proc. 476 : 3K cosmology*, pages 237–+, 1999.
- S. Masi, P. Ade, A. Boscaleri, P. de Bernardis, M. de Petris, G. de Troia, M. Fabbrini, A. Iacoangeli, L. Lamagna, A. Lange, P. Lubin, P. Mauskopf, A. Melchiorri, F. Melchiorri, L. Nati, F. Nati, A. Orlando, F. Piacentini, M. Pierre, G. Pisano, G. Polenta, Y. Rephaeli, G. Romeo, L. Salvaterra, G. Savini, E. Valiante, and D. Yvon. OLIMPO : a balloon-borne, arcminute-resolution survey of the sky at mm and sub-mm wavelengths. In *ESA SP-530 : European Rocket and Balloon Programmes and Related Research*, pages 557–560, August 2003a.
- S. Masi, P. Ade, P. de Bernardis, A. Boscaleri, M. De Petris, G. De Troia, M. Fabbrini, M. Giacometti, A. Iacoangeli, L. Lamagna, A. Lange, P. Lubin, P. Mauskopf, A. Melchiorri, F. Melchiorri, F. Nati, L. Nati, A. Orlando, E. Pascale, F. Piacentini, M. Pierre, G. Polenta, Y. Rephaeli, G. Romeo, and D. Yvon. OLIMPO : A few arcmin resolution survey of the sky at mm and sub-mm wavelengths. *Memorie della Societa Astronomica Italiana*, 74 :96–+, 2003b.
- S. Masi, P. Ade, P. de Bernardis, A. Boscaleri, and et al. M. De Petris. OLIMPO : A few arcmin resolution survey of the sky at mm and sub-mm wavelengths. *Memorie della Societa Astronomica Italiana*, 74 :96, 2003.

- B. S. Mason, T. J. Pearson, A. C. S. Readhead, M. C. Shepherd, J. Sievers, P. S. Udomprasert, J. K. Cartwright, A. J. Farmer, S. Padin, S. T. Myers, J. R. Bond, C. R. Contaldi, U. Pen, S. Prunet, D. Pogosyan, J. E. Carlstrom, J. Kovac, E. M. Leitch, C. Pryke, N. W. Halverson, W. L. Holzapfel, P. Altamirano, L. Bronfman, S. Casassus, J. May, and M. Joy. The Anisotropy of the Microwave Background to $l = 3500$: Deep Field Observations with the Cosmic Background Imager. *ApJ*, 591 :540–555, July 2003. doi : 10.1086/375507.
- B. Mathiesen, A. E. Evrard, and J. J. Mohr. The Effects of Clumping and Substructure on Intracluster Medium Mass Measurements. *ApJ*, 520 :L21–L24, July 1999. doi : 10.1086/312138.
- S. Mei and J. G. Bartlett. On the angular correlation function of SZ clusters : Extracting cosmological information from a 2D catalog. *A&A*, 410 :767–776, November 2003. doi : 10.1051/0004-6361 :20031294.
- A. Melchiorri, P. A. R. Ade, P. de Bernardis, J. J. Bock, J. Borrill, A. Boscaleri, B. P. Crill, G. De Troia, P. Farese, P. G. Ferreira, K. Ganga, G. de Gasperis, M. Giacometti, V. V. Hristov, A. H. Jaffe, A. E. Lange, S. Masi, P. D. Mauskopf, L. Miglio, C. B. Netterfield, E. Pascale, F. Piacentini, G. Romeo, J. E. Ruhl, and N. Vittorio. A Measurement of Ω from the North American Test Flight of Boomerang. *ApJ*, 536 :L63–L66, June 2000. doi : 10.1086/312744.
- J. B. Melin. Amas de galaxies et effet Sunyaev-Zel’dovich : observations et étude des effets de sélection des sondages, 2004.
- J.-B. Melin, J. G. Bartlett, and J. Delabrouille. The selection function of SZ cluster surveys. *A&A*, 429 :417–426, January 2005. doi : 10.1051/0004-6361 :20048093.
- C. J. Miller, C. Genovese, R. C. Nichol, L. Wasserman, A. Connolly, D. Reichart, A. Hopkins, J. Schneider, and A. Moore. Controlling the false-discovery rate in astrophysical data analysis. 112 :3492–3505, dec 2001.
- Y. Moudden, J.-F. Cardoso, J.-L. Starck, and J. Delabrouille. Blind component separation in wavelet space : Application to CMB analysis. *Eurasip Journal on Applied Signal Processing*, 2005. to appear.
- F. Murtagh, J.-L. Starck, and A. Bijaoui. Image restoration with noise suppression using a multiresolution support. *Astronomy and Astrophysics Supplement Series*, 112 :179–189, 1995.
- T. T. Nakamura, T. Matsubara, and Y. Suto. Cosmological redshift distortion : deceleration, bias and density parameters from future redshift surveys of galaxies. *ArXiv Astrophysics e-prints*, June 1997.
- D. M. Neumann, M. Arnaud, S. Majerowicz, and G. Pratt. New results on clusters of galaxies observed with XMM-Newton. In *Two Years of Science with Chandra, Abstracts from the Symposium held in Washington, DC, 5-7 September, 2001, meeting abstract.*, September 2001.

- J. P. Ostriker, P. J. E. Peebles, and A. Yahil. The size and mass of galaxies, and the mass of the universe. *ApJ*, 193 :L1–L4, October 1974.
- B. Paczynski. Gravitational microlensing by the galactic halo. *ApJ*, 304 :1–5, May 1986. doi : 10.1086/164140.
- E. Pantin and J.-L. Starck. Deconvolution of astronomical images using the multiscale maximum entropy method. *Astronomy and Astrophysics Supplement Series*, 315 :575–585, 1996.
- G. Patanchon. Analyse multi-composantes d’observations du fond diffus cosmologique, 2003.
- J. A. Peacock. *Book Review : Cosmological physics / Cambridge U Press, 1999.* 1999.
- P. J. E. Peebles. *The large-scale structure of the universe*. Research supported by the National Science Foundation. Princeton, N.J., Princeton University Press, 1980. 435 p., 1980.
- P. J. E. Peebles and G. Shaviv. The Large-Scale Structure of the Universe. *Space Science Reviews*, 31 :119–+, 1982.
- E. Pierpaoli, S. Anthoine, K. Hufenberger, and I. Daubechies. Reconstructing Sunyaev-Zel’dovich clusters in future cosmic microwave background experiments. *MNRAS*, 359 :261–271, May 2005. doi : 10.1111/j.1365-2966.2005.08896.x.
- S. Pires, J.-B. Juin, Y. Moudén, D. Yvon, S. Anthoine, and E. Pierpaoli. astro-ph/0508641. *submitted to A&A*, 2005.
- E. Pointecouteau, M. Arnaud, J. Kaastra, and J. de Plaa. XMM-Newton observation of the relaxed cluster A478 : Gas and dark matter distribution from $0.01R_{200}$ to $0.5R_{200}$. *A&A*, 423 :33–47, August 2004. doi : 10.1051/0004-6361 :20035856.
- W. H. Press and P. Schechter. Formation of Galaxies and Clusters of Galaxies by Self-Similar Gravitational Condensation. *ApJ*, 187 :425–438, February 1974.
- J.-L. Puget, A. Abergel, J.-P. Bernard, F. Boulanger, W. B. Burton, F.-X. Desert, and D. Hartmann. Tentative detection of a cosmic far-infrared background with COBE. *A&A*, 308 :L5+, April 1996.
- E. D. Reese, J. E. Carlstrom, M. Joy, J. J. Mohr, L. Grego, and W. L. Holzapfel. Determining the Cosmic Distance Scale from Interferometric Measurements of the Sunyaev-Zeldovich Effect. *ApJ*, 581 :53–85, December 2002. doi : 10.1086/344137.
- Y. Rephaeli. Cosmic microwave background comptonization by hot intracluster gas. *ApJ*, 445 :33–36, May 1995. doi : 10.1086/175669.
- Y. Rephaeli. The Sunyaev-Zeldovich effect : Recent progress and future prospects. In *AIP Conf. Proc. 616 : Experimental Cosmology at Millimetre Wavelengths*, pages 309–315, May 2002a.

- Y. Rephaeli. Cosmology with the S-Z Effect. *ArXiv Astrophysics e-prints*, November 2002b.
- Y. Rephaeli. The Sunyaev-Zeldovich Effect. *astro-ph/0511626*, November 2005.
- J. Rich. *Fundamentals of cosmology*. 2002.
- P. Rosati, S. Borgani, and C. Norman. The Evolution of X-ray Clusters of Galaxies. *ARA&A*, 40 :539–577, 2002. doi : 10.1146/annurev.astro.40.120401.150547.
- H. Roussel, R. Sadat, and A. Blanchard. The baryon content of groups and clusters of galaxies. *A&A*, 361 :429–443, September 2000.
- C. L. Sarazin. X-ray emission from clusters of galaxies. *Reviews of Modern Physics*, 58 :1–115, January 1986.
- U. Seljak and M. Zaldarriaga. A Line-of-Sight Integration Approach to Cosmic Microwave Background Anisotropies. *ApJ*, 469 :437–+, October 1996. doi : 10.1086/177793.
- R. K. Sheth, H. J. Mo, and G. Tormen. Ellipsoidal collapse and an improved model for the number and spatial distribution of dark matter haloes. *MNRAS*, 323 :1–12, May 2001. doi : 10.1046/j.1365-8711.2001.04006.x.
- G. F. Smoot, C. L. Bennett, A. Kogut, E. L. Wright, J. Aymon, N. W. Boggess, E. S. Cheng, G. de Amici, S. Gulkis, M. G. Hauser, G. Hinshaw, P. D. Jackson, M. Janssen, E. Kaita, T. Kelsall, P. Keegstra, C. Lineweaver, K. Loewenstein, P. Lubin, J. Mather, S. S. Meyer, S. H. Moseley, T. Murdock, L. Rokke, R. F. Silverberg, L. Tenorio, R. Weiss, and D. T. Wilkinson. Structure in the COBE differential microwave radiometer first-year maps. *ApJ*, 396 :L1–L5, September 1992. doi : 10.1086/186504.
- D. N. Spergel, L. Verde, H. V. Peiris, E. Komatsu, M. R. Nolta, C. L. Bennett, M. Halpern, G. Hinshaw, N. Jarosik, A. Kogut, M. Limon, S. S. Meyer, L. Page, G. S. Tucker, J. L. Weiland, E. Wollack, and E. L. Wright. First-Year Wilkinson Microwave Anisotropy Probe (WMAP) Observations : Determination of Cosmological Parameters. *ApJS*, 148 :175–194, September 2003. doi : 10.1086/377226.
- J.-L. Starck and F. Murtagh. *Astronomical Image and Data Analysis*. Springer-Verlag, 2002.
- J.-L. Starck, F. Murtagh, and A. Bijaoui. *Image Processing and Data Analysis : The Multiscale Approach*. Cambridge University Press, 1998.
- J.-L. Starck, F. Murtagh, P. Querre, and F. Bonnarel. Entropy and astronomical data analysis : Perspectives from multiresolution analysis. *A&A*, 368 :730–746, 2001.
- J.-L. Starck, S. Pires, and A. Refregier. Weak lensing mass reconstruction using fdr. *A&A*, 2005.

- R. A. Sunyaev and I. B. Zeldovich. The velocity of clusters of galaxies relative to the microwave background - The possibility of its measurement. *MNRAS*, 190 : 413–420, February 1980a.
- R. A. Sunyaev and I. B. Zeldovich. Microwave background radiation as a probe of the contemporary structure and history of the universe. *ARA&A*, 18 :537–560, 1980b. doi : 10.1146/annurev.aa.18.090180.002541.
- R. A. Sunyaev and Y. B. Zeldovich. The Interaction of Matter and Radiation in the Hot Model of the Universe. *Ap&SS*, 7 :20–+, 1970.
- R. A. Sunyaev and Y. B. Zeldovich. The Observations of Relic Radiation as a Test of the Nature of X-Ray Radiation from the Clusters of Galaxies. *Comments on Astrophysics and Space Physics*, 4 :173–+, November 1972.
- S. Takahashi and K. Yamashita. XMM-Newton Observation of the Cluster of Galaxies Abell 1650. *PASJ*, 55 :1105–1113, December 2003.
- M. Tegmark, M. A. Strauss, M. R. Blanton, K. Abazajian, S. Dodelson, H. Sandvik, X. Wang, D. H. Weinberg, I. Zehavi, N. A. Bahcall, F. Hoyle, D. Schlegel, R. Scoccimarro, M. S. Vogeley, A. Berlind, T. Budavari, A. Connolly, D. J. Eisenstein, D. Finkbeiner, J. A. Frieman, J. E. Gunn, L. Hui, B. Jain, D. Johnston, S. Kent, H. Lin, R. Nakajima, R. C. Nichol, J. P. Ostriker, A. Pope, R. Scranton, U. Seljak, R. K. Sheth, A. Stebbins, A. S. Szalay, I. Szapudi, Y. Xu, J. Annis, J. Brinkmann, S. Burles, F. J. Castander, I. Csabai, J. Loveday, M. Doi, M. Fukugita, B. Gillespie, G. Hennessy, D. W. Hogg, Ž. Ivezić, G. R. Knapp, D. Q. Lamb, B. C. Lee, R. H. Lupton, T. A. McKay, P. Kunszt, J. A. Munn, L. O’Connell, J. Peoples, J. R. Pier, M. Richmond, C. Rockosi, D. P. Schneider, C. Stoughton, D. L. Tucker, D. E. vanden Berk, B. Yanny, and D. G. York. Cosmological parameters from SDSS and WMAP. *Phys. Rev. D*, 69(10) :103501–+, May 2004. doi : 10.1103/PhysRevD.69.103501.
- J. A. Tyson. Dark matter - Gravitational lens studies. In *AIP Conf. Proc. 222 : After the first three minutes*, pages 437–445, 1991.
- E. A. Valentijn and P. P. van der Werf. First Extragalactic Direct Detection of Large-Scale Molecular Hydrogen in the Disk of NGC 891. *ApJ*, 522 :L29–L33, September 1999. doi : 10.1086/312208.
- M. White, J. Carlstrom, M. Dragovan, and S. W. L. Holzapfel. Analyzing data from DASI. *ArXiv Astrophysics e-prints*, December 1999.
- S. D. M. White, J. F. Navarro, A. E. Evrard, and C. S. Frenk. The Baryon Content of Galaxy Clusters - a Challenge to Cosmological Orthodoxy. *Nature*, 366 :429–+, December 1993.
- Ch. Yèche, Ealet A., Refregier A., Tao C., Tilquin A., Virey J.-M., and Yvon D. Prospects for Dark Energy Evolution : a Frequentist Multi-Probe Approach. *ArXiv Astrophysics e-prints*, July 2005.

- D. Yvon and F. Mayet. Mirage : A new iterative map-making code for CMB experiments. *A&A*, 436 :729–739, June 2005. doi : 10.1051/0004-6361 :20035920.
- Y. B. Zeldovich and R. A. Sunyaev. The Interaction of Matter and Radiation in a Hot-Model Universe. *Ap&SS*, 4 :301–+, 1969.
- F. Zwicky. Die Rotverschiebung von extragalaktischen Nebeln. *Helvetica Physica Acta*, 6 :110–127, 1933.